TRANSACTIONS

OF THE

AMERICAN MATHEMATICAL SOCIETY

EDITED BY

WILLIAM C. GRAUSTEIN

EINAR HILLE

C. C. MACDUFFEE

WITH THE COÖPERATION OF

A. A. ALBERT	ERIC T. BELL	JESSE DOUGLAS
T. H. HILDEBRANDT	E. P. LANE	R. E. LANGER
MARSTON MORSE	OYSTEIN ORE	H. L. RIETZ
H. P. ROBERTSON	M. H. STONE	J. L. SYNGE
GABOR SZEGÖ	G. T. WHYBURN	OSCAR ZARISKI

VOLUME 44

JULY TO DECEMBER, 1938

PUBLISHED BY THE SOCIETY

MENASHA, WIS., AND NEW YORK

1938

Chester C. Corbin Library Fund
Bound Dec. 1936

Composed, Printed and Bound by
The Collegiate Press
George Banta Publishing Company
Menasha, Wisconsin

59584

QA
1
F 001
v.44

TABLE OF CONTENTS

VOLUME 44, JULY TO DECEMBER, 1938

ADKISSON, V. W., of Fayetteville, Ark. Plane peanian continua with unique maps on the sphere and in the plane 58
BAER, R., of Urbana, Ill. Groups with abelian central quotient group 357
BAER, R., of Urbana, Ill. Groups with preassigned central and central quotient group 387
BELL, E. T., of Pasadena, Calif. Polynomial approximations for elliptic functions 47
BLISS, G. A., of Chicago, Ill. Definitely self-adjoint boundary value problems 413
CAMERON, R. H., and MARTIN, W. T., of Cambridge, Mass. Analytic continuation of diagonals and Hadamard compositions of multiple power series 1
CAMP, B. H., of Middletown, Conn. Correction to the paper "The multinomial solid and the chitest" 151
DOOB, J. L., of Urbana, Ill. Stochastic processes with an integral-valued parameter 87
DUNFORD, N., of New Haven, Conn. Uniformity in linear spaces 305
FLOOD, M. M., of Princeton, N. J. Equivalence of pairs of matrices 8
HALL, M., of New Haven, Conn. An isomorphism between linear recurring sequences and algebraic rings 196
HALPERIN, I., of New Haven, Conn. On the transitivity of perspectivity in continuous geometries 537
HAMILTON, O H., of Austin, Texas. Fixed points under transformations of continua which are not connected in Kleinen 18
VAN KAMPEN, E. R., and WINTNER, A., of Baltimore, Md. On the reduction of dynamical systems by means of parametrized invariant relations 168
KASNER, E., of New York, N. Y. The two conformal invariants of fifth order 25
KAWAGUCHI, A., of Sapporo, Japan. Geometry in an n-dimensional space with the arc length $s=\int\{A_i(x, x')x''^i+B(x, x')\}^{1/p}dt$ 153
KLIPPLE, E. C., of College Station, Texas. Two-dimensional spaces in which there exist contiguous points 250
MCSHANE, E. J., of Charlottesville, Va. Some existence theorems in the calculus of variations. I. The Dresden corner condition 429
MCSHANE, E. J., of Charlottesville, Va. Some existence theorems in the calculus of variations. II. Existence theorems for isoperimetric problems in the plane 439

BOSTON UNIVERSITY
COLLEGE OF LIBERAL ARTS
LIBRARY

MADOW, W. G., of New York, N. Y. Contributions to the theory of multivariate statistical analysis 454
MARTIN, W. T., and CAMERON, R. H., of Cambridge, Mass. Analytic continuation of diagonals and Hadamard compositions of multiple power series 1
MORSE, A. P., of Princeton, N. J. A continuous function with no unilateral derivatives 496
OLDENBURGER, R., of Chicago, Ill., and Princeton, N. J. Rational equivalence of a form to a sum of pth powers 219
PETTIS, B. J., of Charlottesville, Va. On integration in vector spaces 277
REID, W. T., of Chicago, Ill. A system of ordinary linear differential equations with two-point boundary conditions 508
SCHOENBERG, I. J., of Waterville, Me. Metric spaces and positive definite functions 522
SYNGE, J. L., of Toronto, Canada. The absolute optical instrument 32
WARD, M., of Pasadena, Calif. The law of apparition of primes in a Lucasian sequence 68
WILLIAMS, W. L., of Chicago, Ill. Permanent configurations in the problem of five bodies 563
WINTNER, A., and VAN KAMPEN, E. R., of Baltimore, Md. On the reduction of dynamical systems by means of parametrized invariant relations 168

ANALYTIC CONTINUATION OF DIAGONALS AND HADAMARD COMPOSITIONS OF MULTIPLE POWER SERIES*

BY

R. H. CAMERON AND W. T. MARTIN

1. **Introduction.** A problem that has always interested mathematicians is the extension of an analytic function defined by a power series outside of its region of convergence. The classical Hadamard theorem† furnishes one method of doing this for functions of one variable. For this theorem gives us information about the analytic continuation of $A(z)=\sum_0^\infty a_n z^n$ provided we can find a way of factoring the coefficients a_n into products $b_n c_n$ such that $B(z)=\sum b_n z^n$ and $C(z)=\sum c_n z^n$ are functions whose analytic continuations are known. Even if we are not able to factor the a_n in such a way, we may be able to find a known function $F(x, y)=\sum f_{mn}x^m y^n$ whose coefficient matrix $\{f_{mn}\}$ has $\{a_n\}$ as its diagonal ($f_{nn}=a_n$). One of the results of the present paper furnishes information concerning the region of analyticity of $A(z)$ in terms of the region of analyticity of $F(x, y)$. In case $F(x, y)=B(x)C(y)$ this method reduces to the preceding.

If we pass now to higher dimensions, our general point of view suggests a number of different possibilities. For example, suppose that we wish to investigate the analytic continuation of $A(x, y)=\sum a_{mn}x^m y^n$. If we can find a known function $F(x_1, x_2, y_1, y_2, y_3)=\sum f_{m_1m_2n_1n_2n_3}x_1^{m_1}x_2^{m_2}y_1^{n_1}y_2^{n_2}y_3^{n_3}$ for which $a_{mn}=f_{mmnnn}$, then simply by iterating our primary theorem we can draw conclusions concerning the analytic extension of A from our knowledge of the analytic extension of F. Similarly we can obtain information concerning $A(x, y, z)=\sum a_{mnp}x^m y^n z^p$ if we can find a known function $F(x_1, x_2, y_1, y_2, z_1, z_2)=\sum f_{m_1m_2n_1n_2p_1p_2}x_1^{m_1}x_2^{m_2}y_1^{n_1}y_2^{n_2}z_1^{p_1}z_2^{p_2}$ for which $a_{mnp}=f_{mmnnpp}$. In fact our basic theorem can be iterated in such a way as to apply to any distribution of dimensions.

Just as our first result contains as a special case the classical Hadamard theorem, so these results contain an n-dimensional generalization of Hadamard's theorem. The regions we consider in the present paper are star-shaped and the nature of the singularities of the functions plays no role whatever.

---

* Presented to the Society, December 29, 1936; received by the editors May 26, 1937.

† J. Hadamard, *Théorème sur les séries entières*, Acta Mathematica, vol. 22 (1898), pp. 55–64.

BOSTON UNIVERSITY
COLLEGE OF LIBERAL ARTS
LIBRARY

Haslam-Jones* has obtained a two-dimensional analogue of the Hadamard theorem which furnishes information about a function $A(x, y)$ if there are two known functions $B(x, y)$, $C(x, y)$ whose coefficients satisfy the relations $b_{mn}c_{mn}=a_{mn}$ and whose singularities satisfy certain conditions. If we compare his theorem with our two-dimensional theorem we see that neither includes the other. For our result does not always furnish as extensive an analytic continuation as that of Haslam-Jones, but it deals with a wider class of functions.

Grouping the terms into homogeneous polymonials, Bochner and Martin† have obtained an n-dimensional analogue of the Hadamard theorem. While (for $n>1$) their theorem does not deal directly with the individual coefficients, it does have the advantage of invariance under any affine transformation.

2. **Stars.** An important class of regions with which we shall be concerned in this paper is the class of star-shaped regions.‡

DEFINITION. A region S in a space R_n of n complex variables $z_1, \cdots, z_n$ is called a $(p_1, \cdots, p_n)$-star if with the point $(z_1, \cdots, z_n)$ it contains also all the points $(\rho^{p_1}z_1, \cdots, \rho^{p_n}z_n)$, where $0 \leqq \rho \leqq 1$ and the p_i are positive integers. In case $p_1=p_2=\cdots=p_n=p$, the $(p_1, \cdots, p_n)$-star is simply called a *star*.

It is clear that only the ratios of the p's are significant. It is also obvious that the union and intersection of two $(p_1, \cdots, p_n)$-stars are also $(p_1, \cdots, p_n)$-stars.

NOTATION. If S is a $(p_1, \cdots, p_n)$-star in R_n, the symbol $S_{(z_1z_2)}$ will denote the point set consisting of all points $P(w, z_3, z_4, \cdots, z_n)$ in R_{n-1} such that to P there corresponds a continuous positive periodic function $r=r(\theta)$ of period 2π for which each of the points $Q(\theta)$: $(re^{i\theta}w, r^{-1}e^{-i\theta}, z_3, \cdots, z_n)$ belongs to S, $(0 \leqq \theta \leqq 2\pi)$.

We shall show that $S_{(z_1z_2)}$ is a $(p_1+p_2, p_3, \cdots, p_n)$-star. Let $P(w, z_3, \cdots, z_n)$ be a point of $S^* \equiv S_{(z_1z_2)}$ and let $r(\theta)$ be the associated function. Then since S is a $(p_1, \cdots, p_n)$-star, each of the points

$$(\rho^{p_1}re^{i\theta}w, \rho^{p_2}r^{-1}e^{-i\theta}, \rho^{p_3}z_3, \cdots, \rho^{p_n}z_n), \qquad (0 \leqq \rho \leqq 1, 0 \leqq \theta \leqq 2\pi),$$

belongs to S. Thus if $R_\rho(\theta)=r(\theta)\rho^{-p_2}$, all of the points

$$(\rho^{p_1+p_2}R_\rho e^{i\theta}w, R_\rho^{-1}e^{-i\theta}, \rho^{p_3}z_3, \cdots, \rho^{p_n}z_n), \qquad (0 < \rho \leqq 1, 0 \leqq \theta \leqq 2\pi),$$

* U. S. Haslam-Jones, *An extension of Hadamard's multiplication theorem*, Proceedings of the London Mathematical Society, (2), vol. 27 (1928), pp. 223–232.

† S. Bochner and W. T. Martin, *Singularities of composite functions in several variables*, Annals of Mathematics, (2), vol. 38 (1937), pp. 293–302.

‡ B. Almer, *Sur quelques problèmes de la théorie des fonctions analytiques de deux variables complexes*, Arkiv för Matematik, Astronomi och Fysik, vol. 17 (1922), pp. 1–70; W. T. Martin, *Special regions of regularity of functions of several complex variables*, Annals of Mathematics, (2), vol. 38 (1937), pp. 602–625.

are in S, and hence all of the points

$$(\rho^{p_1+p_2}w, \rho^{p_2}z_3, \cdots, \rho^{p_n}z^n), \qquad (0 < \rho \leqq 1),$$

are in S^*.

It only remains to show that S^* is an open set, as connectedness obviously follows from the last statement. (It is trivial to show that $(w=0, z_3=0, \cdots, z_n=0)$ belongs to S^* since S contains the origin.) To show that S^* is open, consider a point $P_0(w^{(0)}, z_3^{(0)}, \cdots, z_n^{(0)})$ of it. Since S is open, with $(re^{i\theta}w^{(0)}, r^{-1}e^{-i\theta}, z_3^{(0)}, \cdots, z_n^{(0)})$ there is contained an n-dimensional neighborhood $N(\theta)$ of $Q_0(\theta)$ in S,

$$N(\theta): \quad |z_1 - re^{i\theta}w^{(0)}| < \epsilon(\theta), \quad |z_2 - r^{-1}e^{-i\theta}| < \epsilon(\theta), \quad |z_j - z_j^{(0)}| < \epsilon(\theta), \quad (j = 3, \cdots, n).$$

Because of the continuity of $r(\theta)$ it is clear that there is a positive ϵ independent of θ such that for all θ in $0 \leqq \theta \leqq 2\pi$ all of the points for which $|z_1 - re^{i\theta}w^{(0)}| < \epsilon$, $|z_2 - r^{-1}e^{i\theta}|$, $|z_j - z_j^{(0)}| < \epsilon$, $(j=3, \cdots, n)$, belong to S. Let

$$\delta = \min\left\{\frac{\epsilon}{\max r(\theta)}, \epsilon\right\}.$$

Thus when

$$|w - w^{(0)}| < \delta, \ |z_3 - z_3^{(0)}| < \delta, \cdots, |z_n - z_n^{(0)}| < \delta,$$

the point $(re^{i\theta}w, r^{-1}e^{-i\theta}, z_3, \cdots, z_n)$ is in the ϵ-neighborhood of $Q_0(\theta)$ and hence in S. Thus S^* is a $(p_1+p_2, p_3, \cdots, p_n)$-star. We shall call $S^* \equiv S_{(z_1z_2)}$ a *contracted star* of S. More general contractions will later be given.

DEFINITION. Let $f(z_1, \cdots, z_n)$ be any function analytic in a neighborhood of the origin. Let us define a point set $S \equiv S(f; p_1, \cdots, p_n)$ as follows: A point $(z_1, \cdots, z_n)$ belongs to S if f is regular at all of the points $(\rho^{p_1}z_1, \cdots, \rho^{p_n}z_n)$, $0 \leqq \rho \leqq 1$. It is clear that S is a $(p_1, \cdots, p_n)$-star; we shall call S the $(p_1, \cdots, p_n)$-star of f. We shall denote by $S(f)$ the $(1, \cdots, 1)$-star of f.

3. **Proof of the central theorem.** We shall now prove the following extension of the Hadamard theorem on which our remaining results will be based:

THEOREM 1. *Let*

$$A(z_1, \cdots, z_n) = \sum a_{\mu_1 \cdots \mu_n} z_1^{\mu_1} \cdots z_n^{\mu_n} \tag{1}$$

be analytic in a neighborhood of the origin, and let S be its $(p_1, \cdots, p_n)$-star. Then the diagonal function

$$A_{(z_1z_2)}(w, z_3, \cdots, z_n) = \sum a_{mm\mu_3 \cdots \mu_n} w^m z_3^{\mu_3} \cdots z_n^{\mu_n} \tag{2}$$

is analytic in the contracted star $S^* \equiv S_{(z_1z_2)}$.

For the proof let $P_0(w^{(0)}, z_3^{(0)}, \cdots, z_n^{(0)})$ be any point in S^*. Then by definition of S^* there exists a continuous positive periodic function $r(\theta) = r(\theta; P_0)$ such that all of the points

$$(re^{i\theta}w^{(0)}, r^{-1}e^{-i\theta}, z_3^{(0)}, \cdots, z_n^{(0)}), \qquad 0 \leqq \theta \leqq 2\pi,$$

are in S. Since S is open it is clear that for sufficiently small positive δ all the points $(r^*e^{i\theta}w^{(0)}, r^{*-1}e^{-i\theta}, z_3^{(0)}, \cdots, z_n^{(0)})$ are in S, where

$$r^* = r^*(\theta; P_0) = \frac{1}{\delta}\int_\theta^{\theta+\delta} r(\phi)d\phi.$$

Then the set of points $t = r^*(\theta)e^{i\theta}$ forms a rectifiable closed curve $C^* = C^*(P_0)$ around the origin. For t on C^* and $(w, z_3, \cdots, z_n)$ in a sufficiently small neighborhood $N = N_\epsilon(P_0)$ of P_0,

$$N_\epsilon(P_0): \quad |w - w^{(0)}| < \epsilon, |z_3 - z_3^{(0)}| < \epsilon, \cdots, |z_n - z_n^{(0)}| < \epsilon,$$

the function $A(wt, t^{-1}, z_3, \cdots, z_n)$ is analytic in all its variables. Then since S is a $(p_1, \cdots, p_n)$-star, it follows that $A(u^{p_1}wt, u^{p_2}t^{-1}, u^{p_3}z_3, \cdots, u^{p_n}z_n)$ is analytic in $w, z_3, \cdots, z_n, t, u$ for $P(w, z_3, \cdots, z_n)\epsilon N(P_0)$, $t\epsilon C^*$, $0 \leqq u \leqq 1$. Thus

$$(3) \qquad g(w, z_3, \cdots, z_n, u; P_0) = \frac{1}{2\pi i}\int_{C^*(P_0)} A(u^{p_1}wt, u^{p_2}t^{-1}, u^{p_3}z_3, \cdots, u^{p_n}z_n)\frac{dt}{t}$$

is analytic in $w, z_3, \cdots, z_n, u$ when $P\epsilon N$ and $0 \leqq u \leqq 1$. Now let R be positive and such that the series (1) converges whenever $|z_j| < R$, $(j = 1, \cdots, n)$, and let γ and Γ be the minimum and maximum values of $|t|$ for t on C^*. Let

$$\delta_1 = \min\left\{\left[\frac{R}{(|w_1^{(0)}| + \epsilon)\Gamma}\right]^{1/p_1}, \ (R\gamma)^{1/p_2}, \ \left[\frac{R}{|z_j^{(0)}| + \epsilon}\right]^{1/p_j}\right\},$$

for $j = 3, \cdots, n$. Then if $0 \leqq u < \delta_1$, and $P\epsilon N$,

$$\begin{aligned} g(w, z_3, \cdots, z_n, u; P_0) &= \frac{1}{2\pi i}\int_{C^*}\sum a_{\mu_1\cdots\mu_n}u^{p_1\mu_1}w^{\mu_1}t^{\mu_1}u^{p_2\mu_2}t^{-\mu_2}u^{p_3\mu_3}z_3^{\mu_3}\cdots u^{p_n\mu_n}z_n^{\mu_n}\frac{dt}{t} \\ &= \sum_{m,\mu_3,\cdots,\mu_n=0}^{\infty} a_{m m\mu_3\cdots\mu_n}u^{(p_1+p_2)m+p_3\mu_3+\cdots+p_n\mu_n}w^m z_3^{\mu_3}\cdots z_n^{\mu_n} \\ &= A_{(z_1z_2)}(u^{p_1+p_2}w, u^{p_3}z_3, \cdots, u^{p_n}z_n) \\ &= \sum_{k=0}^{\infty} u^k\left\{\sum_{(p_1+p_2)m+p_3\mu_3+\cdots+p_n\mu_n=k} a_{m m\mu_3\cdots\mu_n}w^m z_3^{\mu_3}\cdots z_n^{\mu_n}\right\} \\ &= \sum_{k=0}^{\infty} B_k(P)u^k. \end{aligned} \tag{4}$$

Since for a fixed $P \varepsilon N$ this is an analytic function of u for $0 \leqq u \leqq 1$, its value at $u=1$ is given by the Le Roy† sum of the last series. Thus

$$(5) \qquad g(P, 1; P_0) = g(w, z_3, \cdots, z_n, 1; P_0) = \lim_{s \to 1^-} \sum_{k=0}^{\infty} \frac{\Gamma(ks+1)}{\Gamma(k+1)} B_k(P).$$

Since $g(P, u; P_0)$ is analytic for $P \varepsilon N$, $0 \leqq u \leqq 1$, $g(P, 1; P_0)$ is certainly analytic for $P \varepsilon N$. But the summation in (5) is independent of P_0, and in particular it is defined when $P = P_0$, while P_0 may be any point in S^*. Thus

$$\tilde{A}(P) = \lim_{s \to 1^-} \sum_{k=0}^{\infty} \frac{\Gamma(ks+1)}{\Gamma(k+1)} B_k(P)$$

is defined throughout S^*; and since $g(P, 1, P_0) = \tilde{A}(P)$ throughout some $N_0(P_0)$, it follows that $\tilde{A}(P)$ is regular at all points of S^*. Since the Le Roy method of summation is regular, it follows that

$$\tilde{A}(P) = \sum_{k=0}^{\infty} B_k(P) = A_{(z_1 z_2)}(P)$$

for P sufficiently close to the origin. Thus our theorem is proved.

4. **Scope of Theorem** 1. As an indication of the scope of Theorem 1, we shall show that it contains the classical Hadamard theorem as a special case. Indeed, if $A(z_1, z_2) = B(z_1)C(z_2)$, where B and C are any two functions analytic in the neighborhood of this origin, then the diagonal function $A_{(z_1 z_2)}$ is the Hadamard composite of B and C. Thus if $\xi_\theta e^{i\theta}$ and $\eta_\theta e^{i\theta}$ are the star vertices of B and C, it is only necessary to show that whenever

$$(6) \qquad 0 \leqq c < \text{g.l.b.}_{0 \leqq \phi \leqq 2\pi} \xi_{\gamma+\phi} \eta_{-\phi}$$

then the point $ce^{i\gamma}$ belongs to $S_{(z_1 z_2)}$, and hence by Theorem 1 is a regular point of $A_{(z_1 z_2)}$. We shall show this by proving the existence of a continuous positive periodic function $r(\theta)$ such that for all θ

$$(7) \qquad c r(\theta) < \xi_{\gamma+\theta}; \qquad \frac{1}{r(\theta)} < \eta_{-\theta}.$$

In the first place it follows from (6) that the set $\mathfrak{B}$ of points $w = te^{i\theta}$ defined by the inequalities

$$(8) \qquad \frac{1}{\eta_{-\theta}} < t < \frac{\xi_{\theta+\gamma}}{c}$$

† See for example, L. L. Smail, *Theory of Summable Infinite Processes*, Oregon, 1925, pp. 13, 14.

contains points on each ray $\arg w = \theta$. Also due to the continuity of the functions z^{-1} and $c^{-1}z$ and the fact that $S(B)$ and $S(C)$ are open regions, the set $\mathfrak{B}$ is an open region. It is obvious that we can find a θ-neighborhood of any θ throughout which there exists a positive continuous function $r_1(\theta)$ which satisfies (7). Thus by the Heine-Borel theorem there is a positive periodic function $r_2(\theta)$ which has at most a finite number of finite jumps in each period interval and which always satisfies (7). Due to (8) we can join the ends of the jumps and obtain a multiple-valued function satisfying (7); and since $\mathfrak{B}$ is open we can modify this function so as to make it single-valued. Thus $r(\theta)$ exists and $ce^{i\gamma}$ belongs to $S_{(z_1z_2)}$.

5. **Generalizations.** Since the contraction of a star is a star, the process of contraction may be iterated. In order to have the resulting stars as extensive as possible, and independent of the order, we make the following definition:

DEFINITION. If S is a $(p_1, \cdots, p_n)$-star, and if

$$T = S_{(z_1z_2)}, \qquad U = S_{(z_3z_4)},$$

then we define

$$S_{(z_1z_2)(z_3z_4)} = T_{(z_3z_4)} + U_{(z_1z_2)},$$

where the sum is a union in the sense of point sets. It is clear that it is a $(p_1+p_2, p_3+p_4, p_5, \cdots, p_n)$-star. Similarly we define

$$S_{(z_1z_2)(z_3z_4)(z_5z_6)} = [S_{(z_1z_2)(z_3z_4)}]_{(z_5z_6)} + [S_{(z_1z_2)(z_5z_6)}]_{(z_3z_4)} + [S_{(z_3z_4)(z_5z_6)}]_{(z_1z_2)}$$

and

$$S_{(z_1z_2z_3)} = T_{(z_{12}z_3)} + U_{(z_{13}z_2)} + V_{(z_{23}z_1)},$$

where

$$T = S_{(z_1z_2)}, \qquad U = S_{(z_1z_3)}, \qquad V = S_{(z_2z_3)}.$$

Contractions of higher orders are defined by the obvious extension of these definitions.

THEOREM 2. *Let* (1) *be analytic in a neighborhood of the origin and let* $S = S(f; p_1, \cdots, p_n)$ *be its* $(p_1, \cdots, p_n)$*-star. Then the diagonal functions*

$$A_{(z_1z_2)(z_3z_4)}(z_{12}, z_{34}, z_5, \cdots, z_n) = \sum a_{mmqq\mu_5\cdots\mu_n} z_{12}^{m} z_{32}^{q} z_5^{\mu_5} \cdots z_n^{\mu_n} \tag{9}$$

and

$$A_{(z_1z_2z_3)}(z_{123}, z_4, \cdots, z_n) = \sum a_{mmm\mu_4\cdots\mu_n} z_{123}^{m} z_4^{\mu_4} \cdots z_n^{\mu_n} \tag{10}$$

are analytic in the stars $S_{(z_1z_2)(z_3z_4)}$ *and* $S_{(z_1z_2z_3)}$ *respectively.*

For the proof we note that

$$A_{(z_1z_2)(z_3z_4)} = [A_{(z_1z_2)}]_{(z_3z_4)} = [A_{(z_3z_4)}]_{(z_1z_2)},$$

and

$$A_{(z_1z_2z_3)} = [A_{(z_1z_2)}]_{(z_{12}z_3)} = [A_{(z_1z_3)}]_{(z_{13}z_2)} = [A_{(z_2z_3)}]_{(z_{23}z_1)}.$$

The theorem follows by iteration of Theorem 1.

6. **Two-dimensional Hadamard theorem.** We can now obtain as a special case of Theorem 2 a two-variable extension of the Hadamard theorem. Let us take a specialized function of four variables; namely, a product of two functions each of two variables,

$$A(s, u, t, v) = B(s, t)C(u, v), \tag{11}$$

$$B(x, y) = \sum b_{mn}x^m y^n, \tag{12}$$

$$C(x, y) = \sum c_{mn}x^m y^n. \tag{13}$$

Then our previously defined diagonal function (see (9)) becomes

$$A_{(su)(tv)}(x, y) = \sum b_{mn}c_{mn}x^m y^n \tag{14}$$

and our special theorem yields the following theorem:

THEOREM 3. *Let* (12) *and* (13) *be any two functions analytic in a neighborhood of the origin, and let* $S(B)$ *and* $S(C)$ *be their stars. Then the Hadamard composition*

$$H(x, y) = \sum b_{mn}c_{mn}x^m y^n$$

is analytic in the star $S^* = S_1 + S_2$, *where* S_1 *and* S_2 *are defined in terms of* $S(B)$ *and* $S(C)$ *as follows. A point* (x, y) *is in the star* S_1 *if for all* ϕ, θ *the points*

$$(xr_\phi(\theta)e^{i\theta}, yr'(\phi)e^{i\theta})$$

are in $S(B)$ *and the points*

$$\left(\frac{1}{r_\phi(\theta)e^{i\theta}}, \frac{1}{r'(\phi)e^{i\theta}}\right)$$

are in $S(C)$, *where* $r'(\phi)$ *is a positive continuous periodic function of* ϕ, *and* $r_\phi(\theta)$ *is a positive periodic continuous function of* θ *and depends also on the parameter* ϕ. *The star* S_2 *is similarly defined, with the roles of* r *and* r' *interchanged.*

From the statement of Theorem 3 and the manner in which it follows from our general theorem, one could easily obtain an n-variable extension of the Hadamard theorem.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY,
CAMBRIDGE, MASS.

EQUIVALENCE OF PAIRS OF MATRICES*

BY

MERRILL M. FLOOD

1. **Introduction.** Two pairs of matrices, $[A_1, A_2]$ and $[B_1, B_2]$, with elements in a commutative field F, are said to be *equivalent*† if and only if there exist two non-singular matrices P and Q, with elements in F, such that $A_1 = PB_1Q$ and $A_2 = PB_2Q$.

The totality of pairs of matrices may be separated into different classes in such a way that all pairs in one class are equivalent to one another while pairs in different classes are not equivalent. The problem which arises naturally is to determine a set of invariants which will characterize the pairs in each class and to select from each class a *canonical* pair defined uniquely in terms of these invariants. A *rational* solution of the problem is one which is carried out completely in the field F; the invariants and canonical pairs obtained in such a solution will be *rational*.

The *rank* of a pair $[A_1, A_2]$ is the maximum rank of the matrices of the *matric pencil* $A = A_1x_1 + A_2x_2$, where x_1 and x_2 are indeterminates in F. A matric pencil is said to be *non-singular* if it is square and of rank equal to its order (otherwise it is called *singular*), and it is said to be *regular* if the rank of either one of its coefficients is the same as the rank of the pencil.

Non-singular matric pencils were first classified by Weierstrass‡ who constructed an irrational canonical form defined by means of the elementary divisors of the pencil. Frobenius§ later gave a rational treatment of the non-singular case. Kronecker‖ treated the singular case and gave an irrational canonical form. Muth¶ gave a full account of the theory of pairs of bilinear forms as it stood at the turn of the century. De Séguier** seems to have been the first to give a rational treatment of the singular case. More recently, it has received the attention of Dickson,†† Turnbull and Aitken,‡‡ Wedderburn,§§

* Presented to the Society, March 27, 1937; received by the editors May 25, 1937.

† C. C. MacDuffee, *The Theory of Matrices*, Berlin, 1933, p. 48.

‡ K. Weierstrass, Monatsberichte, Preussische Akademie der Wissenschaften, 1868, pp. 310–338.

§ G. Frobenius, Sitzungsberichte, Preussische Akademie der Wissenschaften, 1894, pp. 31–44.

‖ L. Kronecker, Sitzungsberichte, Preussische Akademie der Wissenschaften, 1890, pp. 1225–1237.

¶ P. Muth, *Theorie und Anwendung der Elementartheiler*, Leipzig, 1899.

** J. A. de Séguier, Bulletin de la Société Mathématique de France, vol. 36 (1908), pp. 20–40.

†† L. E. Dickson, these Transactions, vol. 29 (1927), pp. 239–253.

‡‡ Turnbull and Aitken, *Canonical Matrices*, Glasgow, 1932, chap. 9.

§§ J. H. M. Wedderburn, *Lectures on Matrices*, American Mathematical Society Colloquium Publications, vol. 17, 1934, chap. 4.

Turnbull,* Ledermann,† Williamson,‡ and others.

In this paper the problem of constructing a rational canonical form in the singular case is reduced to the consideration of the non-singular case. The proofs are completely rational, quite elementary, and relatively short. The canonical form which is obtained is defined essentially in terms of the set of invariants shown by Williamson‡ to characterize the classes of equivalent matrices. The method of proof is very similar to that used by Ingraham§ in his treatment of the equivalence of singular pencils of Hermitian matrices.

2. **Preliminary remarks.** Consider a singular pencil $A=A_1x_1+A_2x_2$ of rank $\rho(A)=r$ and order $[\theta, \theta']$. Set $R_1=A_1t_{11}+A_2t_{21}$ and $R_2=A_1t_{12}+A_2t_{22}$, where the t_{ij} are quantities of F such that $|t_{ij}| \neq 0$. If $R=R_1x_1+R_2x_2$ and $T=\|t_{ij}\|$, then the relations above may be written $R=AT$, and the pencils A and R are said to be *transformable*. If B is a second matric pencil, it follows easily that A is equivalent to B $(A\sim B)$ if and only if $AT\sim BT$. In particular, there exist two quantities t_{11}^0 and t_{21}^0 of F not both zero and such that $\rho(A_1t_{11}^0+A_2t_{21}^0)=r$, and in this case the pencil $R=AT$ is said to be regular.

If it is desired only to obtain necessary and sufficient conditions for the equivalence of two pencils, then there is no loss of generality in considering only regular pencils. However, if a canonical form in the most strict sense is required, it is necessary to start with the original pencils rather than their regular transforms, as has been pointed out by Ledermann.‖ Canonical forms will be constructed only for regular pencils, but the invariants used will be shown to afford a satisfactory classification for all pencils. It is felt that this solves the important part of the problem.

3. **Rational canonical form for regular matric pencils.** Constant non-singular matrices P and Q exist¶ such that

$$PR_1Q = e = \left\| \begin{array}{cc} 1_r & 0 \\ 0 & 0 \end{array} \right\|;$$

hence $R\sim ex_1+PR_2Qx_2=ex_1+ax_2=R_0$. If we set

$$R_0 = \left\| \begin{array}{cc} 1_r & 0 \\ 0 & 0 \end{array} \right\| x_1 + \left\| \begin{array}{cc} a_{11} & a_{12} \\ a_{21} & a_{22} \end{array} \right\| x_2,$$

---

* H. W. Turnbull, Proceedings of the Edinburgh Mathematical Society, (2), vol. 4 (1935) pp. 67–76.

† W. Ledermann, ibid., (2), vol. 4 (1935), pp. 92–105.

‡ J. Williamson, ibid., (2), vol. 4 (1936), pp. 224–231.

§ Ingraham and Wegner, these Transactions, vol. 38 (1935), pp. 145–162.

‖ Loc. cit.

¶ MacDuffee, loc. cit., p. 43.

it follows immediately that $a_{22}=0$, for otherwise the rank of R_0 would be greater than r, which is impossible since R_0 has been assumed to be regular.

Since $1_r x_1+a_{11}x_2$ is non-singular, the rows of $(a_{21}\ 0)$ must be linearly dependent on the rows of $(1_r x+a_{11}\ a_{12})$; thus there exists a matrix X_{21} such that $X_{21}(1_r x+a_{11})=a_{21}$ and $X_{21}a_{12}=0$. Necessarily then, $X_{21}=a_{21}(1_r x+a_{11})^{-1}$, and $a_{21}(1_r x+a_{11})^{-1}a_{12}=0$. For x sufficiently large

$$(1_r x+a_{11})^{-1}=\frac{1}{x}\sum_{k=0}^{\infty}\left(\frac{-a_{11}}{x}\right)^k;$$

hence $a_{21}a_{11}^k a_{12}=0$ for $k=0, 1, 2, \cdots$.

Conversely, if $a_{21}a_{11}^k a_{12}=0$ for $k=0, 1, 2, \cdots$, then $a_{21}(1_r x+a_{11})^{-1}a_{12}=0$. Hence, if $X_{21}=a_{21}(1_r x+a_{11})^{-1}$, it follows that $X_{21}(1_r x+a_{11})=a_{21}$ and $X_{21}a_{12}=0$, so that the last $\theta-r$ rows of $ex+a$ are dependent on the first r rows. This proves the following lemma:

LEMMA A. *The rank of the matric pencil*

$$\left\|\begin{matrix}1_r & 0\\ 0 & 0\end{matrix}\right\| x_1+\left\|\begin{matrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{matrix}\right\| x_2$$

is r if and only if $a_{22}=0$ and $a_{21}a_{11}^k a_{12}=0$ for $k=0, 1, 2, \cdots$.

Lemma A holds true only if the coefficient field F is commutative and has characteristic zero. For a field of characteristic $p\neq 0$ it is necessary to alter the treatment slightly. The present author has shown (Annals of Mathematics, (2), vol. 36 (1935), p. 865) that the matric pencil of Lemma A is equivalent to a pencil

$$\left\|\begin{matrix}1_r & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 0\end{matrix}\right\| x_1+\left\|\begin{matrix}b_{11} & 0 & b_{13}\\ 0 & 1_s & 0\\ b_{31} & 0 & 0\end{matrix}\right\| x_2,$$

where $s=\rho(a_{22})$ and $b_{31}b_{11}^k b_{13}=0$ for $k=0, 1, 2, \cdots$. It follows easily that there is no loss of generality in considering pencils satisfying the conditions of Lemma A; and in this way the proofs given in this paper may be extended to include pencils with coefficients in an arbitrary field. This more general method of proof has been used by the present author in a recent paper (*Strict equivalence of matric pencils*, presented to the Society December 29, 1937, but not yet published) treating the problem of equivalence of matric pencils, singular and non-singular.

We now proceed with the construction of a canonical form for the regular pencil $S=ex+a$. If α and β are non-singular matrices of orders θ and θ' with elements in F, then $\alpha e\beta=e$ if

$$\alpha = \left\| \begin{matrix} \alpha_{11} & \alpha_{12} \\ 0 & \alpha_{22} \end{matrix} \right\| \quad \text{and} \quad \beta = \left\| \begin{matrix} \alpha_{11}^{-1} & 0 \\ \beta_{21} & \beta_{22} \end{matrix} \right\|,$$

where α_{22} and β_{22} are non-singular, and so

$$\alpha a \beta = \left\| \begin{matrix} \alpha_{11}a_{11}\alpha_{11}^{-1} + \alpha_{12}a_{21}\alpha_{11}^{-1} + \alpha_{11}a_{12}\beta_{21} & \alpha_{11}a_{12}\beta_{22} \\ \alpha_{22}a_{21}\alpha_{11}^{-1} & 0 \end{matrix} \right\|.$$

Now if the rank of a_{21} is r_1, α_{22} and α_{11} may be chosen so that

$$\alpha_{22}a_{21}\alpha_{11}^{-1} = \left\| \begin{matrix} 1_{r_1} & 0 \\ 0 & 0 \end{matrix} \right\|,$$

and it follows from Lemma A that the first r_1 rows of $\alpha_{11}a_{12}\beta_{22}$ must be zero. Furthermore, if the rank of a_{12} is c_1, it is clear that β_{22} may be chosen so that

$$\alpha_{11}a_{12}\beta_{22} = \left\| \begin{matrix} 0 & 0 \\ 0 & 1_{c_1} \end{matrix} \right\|.$$

With this choice of α_{ij} and β_{kk} the pencil $T = \alpha S \beta$ takes the form

$$T = \left\| \begin{matrix} 1_{r_1} & 0 & 0 & 0 & 0 & 0 \\ 0 & 1_h & 0 & 0 & 0 & 0 \\ 0 & 0 & 1_{c_1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0_{r_1} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0_{jk} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0_{c_1} \end{matrix} \right\| x + \left\| \begin{matrix} x_1 & a_1 & a_4 & 0 & 0 & 0 \\ x_2 & a_2 & a_3 & 0 & 0 & 0 \\ x_3 & x_4 & x_5 & 0 & 0 & 1_{c_1} \\ 1_{r_1} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{matrix} \right\|,$$

where $h = r - r_1 - c_1$, $j = \theta - r - r_1 - c_1$, and $k = \theta' - r - r_1 - c_1$. Finally α_{12} and β_{21} may clearly be chosen so that the x_i are all zero; then

$$T = \left\| \begin{matrix} x & a_1 & a_4 & 0 & 0 & 0 \\ 0 & x + a_2 & a_3 & 0 & 0 & 0 \\ 0 & 0 & x & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{matrix} \right\|.$$

The rank of T is r, since α and β were chosen non-singular, therefore

$$\left\| \begin{matrix} a_1 & a_4 \\ x + a_2 & a_3 \end{matrix} \right\|$$

is regular and of rank $r-r_1-c_1$. Hence, by Lemma A, $a_4=0$ and $a_1a_2^k a_3=0$ for $k=0, 1, 2, \cdots$.

Consider a second regular pencil $U=U_1x_1+U_2x_2$ which is equivalent to R. The rank of U_1 is necessarily r, therefore, constant non-singular matrices P_0 and Q_0 exist such that $P_0U_1Q_0=e$. Then if

$$V = P_0UQ_0 = ex_1 + bx_2 = ex_1 + \left\| \begin{matrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{matrix} \right\| x_2,$$

it follows from Lemma A that $b_{22}=0$. Since $V\sim S$ there exist constant non-singular matrices x and y such that $xV=Sy$, which equation is equivalent to the relations $xe=ey$ and $xb=ay$. From $xe=ey$ it follows that $x_{11}=y_{11}$, $x_{21}=0$, $y_{12}=0$, and that x_{11}, x_{22}, and y_{22} are non-singular. Then from $xb=ay$ it follows that $x_{22}b_{21}=a_{21}x_{11}$ and $x_{11}b_{12}=a_{12}y_{22}$. Since x_{22}, x_{11}, and y_{22} are necessarily non-singular this shows that the ranks of b_{21} and b_{12} are the same as the ranks r_1 and c_1 of a_{21} and a_{12}. r_1 will be called the first "row invariant subrank" and c_1 the first "column invariant subrank" of R or of any pencil equivalent to R. It follows that constant non-singular matrices α_0 and β_0 can be chosen so that $W=\alpha_0V\beta_0$ takes a form analogous to that of T but with a_k replaced by b_k.

The pencils U and R are equivalent if and only if there exist constant non-singular matrices p and q such that $pT=Wq$. From $pe=eq$ it is clear that p and q must be of the forms

$$p = \left\| \begin{matrix} z & * \\ 0 & * \end{matrix} \right\| \quad \text{and} \quad q = \left\| \begin{matrix} z & 0 \\ * & * \end{matrix} \right\|,$$

hence $pT=Wq$ may be replaced by the equation

$$\left\| \begin{matrix} p_{11} & p_{12} & p_{13} & p_{14} & p_{15} & p_{16} \\ p_{21} & p_{22} & p_{23} & p_{24} & p_{25} & p_{26} \\ p_{31} & p_{32} & p_{33} & p_{34} & p_{35} & p_{36} \\ 0 & 0 & 0 & p_{44} & p_{45} & p_{46} \\ 0 & 0 & 0 & p_{54} & p_{55} & p_{56} \\ 0 & 0 & 0 & p_{64} & p_{65} & p_{66} \end{matrix} \right\| \left\| \begin{matrix} 0 & a_1 & 0 & 0 & 0 & 0 \\ 0 & a_2 & a_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{matrix} \right\| = \left\| \begin{matrix} 0 & b_1 & 0 & 0 & 0 & 0 \\ 0 & b_2 & b_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{matrix} \right\| \left\| \begin{matrix} q_{11} & q_{12} & q_{13} & 0 & 0 & 0 \\ q_{21} & q_{22} & q_{23} & 0 & 0 & 0 \\ q_{31} & q_{32} & q_{33} & 0 & 0 & 0 \\ q_{41} & q_{42} & q_{43} & q_{44} & q_{45} & q_{46} \\ q_{51} & q_{52} & q_{53} & q_{54} & q_{55} & q_{56} \\ q_{61} & q_{62} & q_{63} & q_{64} & q_{65} & q_{66} \end{matrix} \right\|.$$

This equation is equivalent to the set of relations:

$$\begin{aligned}
&p_{54} = p_{64} = p_{12} = p_{13} = q_{64} = q_{65} = p_{23} = 0, \\
&p_{11} = p_{44}, \quad p_{33} = q_{66}, \quad p_{34} = q_{61}, \\
&p_{14} = b_1p_{21}, \qquad p_{32}a_3 = q_{63}, \qquad p_{12}a_3 = b_1p_{23}, \\
&p_{24} = b_2p_{21} + b_3p_{31}, \qquad p_{11}a_1 + p_{12}a_2 = b_1p_{22}, \\
&p_{31}a_1 + p_{32}a_2 = q_{62}, \qquad p_{22}a_3 = b_2p_{23} + b_3p_{33}, \\
&p_{21}a_1 + p_{22}a_2 = b_2p_{22} + b_3p_{32}.
\end{aligned}$$

It follows easily that these equations have a solution such that p is non-singular if and only if there exist constant non-singular matrices p_{11}, p_{22}, and p_{33} which satisfy the relations:

$$p_{11}a_1 = b_1p_{22}, \qquad p_{22}a_3 = b_3p_{33},$$
$$p_{21}a_1 + p_{22}a_2 = b_2p_{22} + b_3p_{32}.$$

These equations may be rewritten in the form

$$\begin{Vmatrix} p_{22} & p_{21} \\ 0 & p_{11} \end{Vmatrix} \begin{Vmatrix} a_2 & a_3 \\ a_1 & 0 \end{Vmatrix} = \begin{Vmatrix} b_2 & b_3 \\ b_1 & 0 \end{Vmatrix} \begin{Vmatrix} p_{22} & 0 \\ p_{32} & p_{33} \end{Vmatrix},$$

and this is simply the condition for the equivalence of the two regular pencils

$$T^1 = \begin{Vmatrix} 1 & 0 \\ 0 & 0 \end{Vmatrix} x + \begin{Vmatrix} a_2 & a_3 \\ a_1 & 0 \end{Vmatrix}, \qquad W^1 = \begin{Vmatrix} 1 & 0 \\ 0 & 0 \end{Vmatrix} x + \begin{Vmatrix} b_2 & b_3 \\ b_1 & 0 \end{Vmatrix}.$$

The pencils T^1 and W^1 will be called "first kernels" of the pencils R and U. Thus the problem of classifying singular pencils of rank r has been reduced to that of classifying singular pencils of rank $r-r_1-c_1$, or else to that of classifying non-singular pencils if a_1 and a_3 happen to be zero.

If r_{j+1} and c_{j+1} are the first invariant subranks of T^j, and T^{j+1} is a first kernel of T^j for $j=1, 2, 3, \cdots, n$, and if T^{n+1} is non-singular or zero; then r_j and c_j for $j=1, 2, \cdots, n+1$ will be called the "invariant subranks" of R, and T^{n+1} a "kernel" of R. This proves the following:

THEOREM 1. *Two regular matric pencils are equivalent if and only if they have identical sets of invariant subranks and equivalent kernels.*

It is clear that the construction which leads to T can be extended until a rational canonical form for R is obtained. This canonical form would display the invariant subranks and invariant factors of R. The invariant factors of R are clearly the same as those of any kernel of R except for $\sum_{j=1}^{n+1}(r_j+c_j)$ units which would appear in the normal form of R but not in the normal form of any kernel of R. This demonstrates the corollary:

COROLLARY 1.1. *Two regular matric pencils are equivalent if and only if they have the same invariant subranks and the same elementary divisors.*

4. **Transformable matric pencils.** Let $A = A_1x_1+A_2x_2$ be an arbitrary matric pencil, and define matrices $M_k(A)$ and $N_k(A)$, for $k=1\ 2, 3, \cdots$, by the relations

$$M_1(A) = \| A_1 \quad A_2 \|, \qquad M_2(A) = \begin{Vmatrix} A_1 & A_2 & 0 \\ 0 & A_1 & A_2 \end{Vmatrix}, \cdots,$$

$$N_1(A) = \left\| \begin{matrix} A_1 \\ A_2 \end{matrix} \right\|, \qquad N_2(A) = \left\| \begin{matrix} A_1 & 0 \\ A_2 & A_1 \\ 0 & A_2 \end{matrix} \right\|, \cdots .$$

It is convenient to denote by $m_k(A)$ and $n_k(A)$ the ranks of $M_k(A)$ and $N_k(A)$ and to call $m_k(A)$ the "row singularities" and $n_k(A)$ the "column singularities" of A.

It is obvious that equivalent matric pencils have the same row and column singularities. We now proceed to prove the following:

THEOREM 2. *Transformable matric pencils have the same singularities.*

Proof.* Consider a matric pencil $A = A_1x_1 + A_2x_2$ and a non-singular transformation of indeterminates $x = tx'$, or more explicitly

$$x_1 = t_{11}x_1' + t_{12}x_2', \qquad x_2 = t_{21}x_1' + t_{22}x_2'.$$

Under this transformation, the pencil A is carried into the pencil

$$A' = At = (t_{11}A_1 + t_{21}A_2)x_1' + (t_{12}A_1 + t_{22}A_2)x_2' = A_1'x_1' + A_2'x_2',$$

and the theorem states that $m_k(A) = m_k(A')$ and $n_k(A) = n_k(A')$. The first of these equalities will be demonstrated by constructing non-singular matrices T^k such that

$$M_k(A)T^k = T^{k-1}M_k(A') \qquad \text{for} \qquad k = 1, 2, 3, \cdots,$$

and the second can be shown by a similar construction.

If $u_0, u_1, u_2, \cdots, u_k$ are $k+1$ indeterminates, then the identity

$$\begin{aligned} (1) \qquad & u_0x_1^k + C_{k,1}u_1x_1^{k-1}x_2 + C_{k,2}u_2x_1^{k-2}x_2^2 + \cdots + u_kx_2^k \\ & \equiv u_0'x_1'^k + C_{k,1}u_1'x_1'^{\,-1}x_2' + C_{k,2}u_2'x_1'^{k-2}x_2'^2 + \cdots + u_k'x_2'^k \end{aligned}$$

defines $u_0', u_1', \cdots, u_k'$ as linear combinations of $u_0, u_1, \cdots, u_k$, and these may be written in either of the forms

$$u_i' = \sum_{j=0}^{k} T_{ij}^k u_j \qquad \text{or} \qquad u' = T^k u.$$

Now if $T^k = \| T_{ij}^k \|$, then $t \to T^k$ is a representation of the full linear group of all non-singular matrices of order two, and hence T^k is non-singular since t is.

If (1) is differentiated with respect to x_1' and x_2', there results:

$$\begin{aligned} (2a) \qquad & (u_0t_{11}+u_1t_{21})x_1^{k-1} + C_{k-1,1}(u_1t_{11}+u_2t_{21})x_1^{k-2}x_2 + \cdots + (u_{k-1}t_{11}+u_kt_{21})x_2^{k-1} \\ & \equiv u_0'x_1'^{k-1} + C_{k-1,1}u_1'x_1'^{\,-2}x_2' + \cdots + u_{k-1}'x_2'^{k-1}, \end{aligned}$$

* I am indebted to Dr. A. H. Clifford for this proof.

$$(2b)\quad \begin{aligned}&(u_0 t_{12}+u_1 t_{22})x_1^{k-1}+C_{k-1,1}(u_1 t_{12}+u_2 t_{22})x_1^{k-2}x_2+\cdots+(u_{k-1}t_{12}+u_k t_{22})x_2^{k-1}\\ &\equiv u_1' x_1'^{k-1}+C_{k-1,1}u_2' x_1'^{k-2}x_2'+\cdots+u_k' x_2'^{k-1}.\end{aligned}$$

If (2a) is multiplied by A_1, and (2b) by A_2, and the resulting equations are added, it follows that

$$\begin{aligned}&(u_0A_1'+u_1A_2')x_1^{k-1}+C_{k-1,1}(u_1A_1'+u_2A_2')x_1^{k-2}x_2+\cdots+(u_{k-1}'A_1'+u_kA_2')x_2^{k-1}\\ &\equiv(u_0'A_1+u_1'A_2)x_1'^{k-1}+C_{k-1,1}(u_1'A_1+u_2'A_2)x_1'^{k-2}x_2'+\cdots+(u_{k-1}'A_1+u_2'A_2)x_2'^{k-1},\end{aligned}$$

and from this identity that

$$u_i'A_1+u_{i+1}'A_2\equiv\sum_{j=0}^{k-1}T_{ij}^{k-1}(u_jA_1'+u_{j+1}A_2')\quad\text{for}\quad i=0,1,\cdots,k-1.$$

These identities may be written in the form

$$M_k(A)u'\equiv T^{k-1}M_k(A')u$$

or, since $u'=T^ku$, in the form

$$M_k(A)T^ku\equiv T^{k-1}M_k(A')u.$$

The indeterminate vector u may be cancelled in this equation and so

$$M_k(A)T^k=T^{k-1}M_k(A')$$

as was to be shown.

It is convenient, at this point, to state the following:

LEMMA B.* *The invariant factors of transformable matric pencils are connected by the same transformation of the indeterminates x_1 and x_2 as the pencils themselves.*

5. **Equivalence of general matric pencils.** Williamson† has shown that the minimal numbers of a matric pencil can be expressed in terms of its singularities, from which follows the theorem:

THEOREM 3. *Two matric pencils are equivalent if and only if they have the same singularities and invariant factors.*

This theorem may also be proved with the help of Theorem 2 and Lemma B by showing that the invariant subranks of a regular matric pencil can be expressed in terms of its singularities. This will now be done for the row subranks, and an analogous treatment of the column subranks would complete the proof. There is no loss of generality if the pencil is taken to be in canonical form.

* See MacDuffee, loc. cit.

† Loc. cit.

Consider the regular canonical pencil $W=e_0x+a$ of rank r, with row subranks r_k and column subranks c_k. Set

$$E_k = \left\| \begin{matrix} 1_{r_k} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix} \right\|, \qquad F_k = \left\| \begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1_{c_k} \end{matrix} \right\|, \quad \text{and} \quad a = \left\| \begin{matrix} E_1 & 0 \\ A_1 & F_1 \end{matrix} \right\|.$$

Then square matrices A_k may be defined by the relations

$$A_k = \left\| \begin{matrix} 0 & E_k & 0 \\ 0 & A_k & F_k \\ 0 & 0 & 0 \end{matrix} \right\| \qquad \text{for} \qquad k = 1, 2, \cdots, \phi,$$

where $A_{\phi+1}$ is the canonical kernel of S and $E_kA_k^jF_k=0$ for $j=1, 2, \cdots$. Of course r_k is the rank of E_k, and c_k is the rank of F_k.

Now, by definition,

$$m_2(W) = \rho \left\| \begin{matrix} e_0 & a & 0 \\ 0 & e_0 & a \end{matrix} \right\| = \rho \left\| \begin{matrix} 0 & 0 & E_1 & 0 & 0 & 0 \\ 1 & 0 & A_1 & F_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & E_1 & 0 \\ 0 & 0 & 1 & 0 & A_1 & F_1 \end{matrix} \right\| = r + \rho \left\| \begin{matrix} E_1 & 0 & 0 \\ 0 & E_1 & 0 \\ 1 & A_1 & F_1 \end{matrix} \right\|.$$

If the first column of this matrix is multiplied by $-F_1$ and added to the last column, and then the third row is multiplied by $-E_1$ and added to the first row; since $E_1F_1=0$, it follows that

$$m_2(W) = r + \rho \left\| \begin{matrix} 0 & -E_1A_1 \\ 0 & E_1 \\ 1 & A_1 \end{matrix} \right\| = 2r + \rho \left\| \begin{matrix} E_1 \\ E_1A_1 \end{matrix} \right\|.$$

In similar fashion, it is easily shown that

$$m_k(W) = kr + \rho \left\| \begin{matrix} E_1 \\ E_1A_1 \\ E_1A_1^2 \\ \cdots \\ E_1A_1^{k-1} \end{matrix} \right\| \qquad \text{for} \quad k = 1, 2, 3, \cdots.$$

Since $E_2A_2^jF_2=0$, it follows that

$$A_1^j = \left\| \begin{matrix} 0 & E_2A_2^{j-1} & 0 \\ 0 & A_2^j & A_2^{j-1}F_2 \\ 0 & 0 & 0 \end{matrix} \right\|,$$

and hence that

$$\rho \left\| \begin{matrix} E_1 \\ E_1A_1 \\ \cdot \cdot \cdot \\ E_1A_1^{k-1} \end{matrix} \right\| = r_1 + \rho \left\| \begin{matrix} E_2 \\ E_2A_2 \\ \cdot \cdot \cdot \\ E_2A_2^{k-1} \end{matrix} \right\|;$$

and a simple induction now shows that

$$m_k(W) = kr + \sum_{j=1}^{k-1} r_j.$$

This equation provides the necessary relationship between the row singularities and subranks of W and the proof of Theorem 3 is complete. Of course,

$$r_k = m_k - r - \sum_{j=1}^{k-1} m_j \qquad \text{and} \qquad c_k = n_k - r - \sum_{j=1}^{k-1} n_j$$

are the inverse equations which express the subranks in terms of the singularities.

PRINCETON UNIVERSITY,
PRINCETON, N. J.

FIXED POINTS UNDER TRANSFORMATIONS OF CONTINUA WHICH ARE NOT CONNECTED IM KLEINEN*

BY

O. H. HAMILTON

1. **Introduction.** Ayres† has shown that if M is a compact continuous curve in the plane which does not separate the plane and T is a reversibly continuous transformation of M into a subset of itself, then T leaves some point of M invariant. He also proves similar theorems for special cases of non-planar continuous curves. The purpose of this paper is to extend these results to certain types of continua which are not connected im kleinen.

2. **Preliminary theorem and lemmas.** We prove first the following preliminary theorems and lemmas.

THEOREM I. *If M is a compact continuum in a metric space but is not an indecomposable continuum, and if T is any reversibly continuous transformation of M into a subset of itself, then some proper subcontinuum N of M contains a point of $T(N)$, the image of N under T.*

Proof. If T carries M into a proper subset of itself, then $T(M)$ contains its image $T^2(M)$ and the theorem is true. Suppose then that T carries M into itself. It can easily be shown that if M is not indecomposable, it does not contain two mutually exclusive composants. The transformation T carries a composant of M into a composant of M. It has been shown that a composant‡ of a continuum is the sum of a countable number of continua $N_1, N_2, N_3, \cdots$, where for each integer i, N_i contains N_{i-1}.

Let V be any composant of M, and let $T(V)$ be its image under T. Then V and $T(V)$, by what was said above, must have a point in common. Let V be expressed as $\sum_{i=1}^{\infty} V_i$, where for each integer i, V_i is a continuum which contains V_{i-1}. Then $T(V) = \sum_{i=1}^{\infty} T(V_i)$ where for each integer i, $T(V_i)$ is the image of V_i under T and is a continuum which contains $T(V_{i-1})$. For some integer j, V_j contains a point of $T(V_j)$, for if we suppose the contrary to be true, it is obvious that V contains no point of $T(V)$, contradicting the fact

---

* Presented to the Society, September 10, 1937; received by the editors May 29, 1937.

† W. L. Ayres, *Some generalizations of the Scherrer fixed-point theorem*, Fundamenta Mathematicae, vol. 16 (1930), pp. 333–336.

‡ See R. L. Moore, *Foundations of Point Set Theory*, American Mathematical Society Colloquium Publications, vol. 13, p. 75.

that V and $T(V)$ have a point in common. V_i is therefore a continuum satisfying the conclusions of the theorem.

LEMMA 1. *If M is a compact continuum in a metric space and M is not an indecomposable continuum and contains no continuum which is the sum of two continua whose common part is disconnected, and if T is a reversibly continuous transformation of M into a subset of itself and neither T nor T^{-1} carries a proper subcontinuum of M into itself, then there is a proper subcontinuum L_0 of M such that L_0 contains a point of $T(L_0)$ but such that for $n>1$, the common part of L_0 and $T^n(L_0)$ is vacuous.*

Proof. By Theorem I, some proper subcontinuum K_0 of M contains a point of its image $T(K_0)$. For each positive or negative integer n, let K_n designate $T^n(K_0)$. By a hypothesis of the theorem, K_0 is not identical with K_1, for suppose $K_0=K_1$. Then K_0 is a proper subcontinuum of M which is carried into itself by T. Furthermore $K_0 \cdot K_1$ is not identical with K_0 or with K_1; for suppose $K_0 \cdot K_1$ is identical with K_0. Then T carries K_0 into a subset of itself. Consider the continuum $A=\prod_{n=0}^{\infty} K_n$. Then $T(A)=A$, and T carries A, a proper subcontinuum of M, into itself. Similarly, if $K_0 \cdot K_1$ is identical with K_1, T^{-1} carries a proper subcontinuum of M into itself; thus we have a contradiction of a hypothesis of the theorem.

There exists an integer r such that the $\prod_{i=0}^{r} K_i$ is not vacuous but such that $\prod_{i=0}^{r+1} K_i$ is vacuous. This can be shown as follows. $K_0 \cdot K_1$ is not vacuous. Suppose $\prod_{i=0}^{s} K_i$ contains a point for each integer s. Then $B=\prod_{n=0}^{\infty} K_n$ is not vacuous and is a proper subcontinuum of M which is carried into itself by T, which is, again, a contradiction of a hypothesis of the theorem. Let L_0 be the continuum $\prod_{i=0}^{r-1} K_i$, where r is an integer such that $\prod_{i=0}^{r} K_i$ is not vacuous but $\prod_{i=0}^{r+1} K_i$ is vacuous. For each integer i, let $L_i=T^i(L_0)$. Then L_1 is $\prod_{i=1}^{r} K_i$. $L_0 \cdot L_1$ is not vacuous since $L_0 \cdot L_1=\prod_{i=0}^{r} K_i$ which is not vacuous. $L_0 \cdot L_2$ is vacuous. For suppose L_0 contains a point of L_2. Then $\prod_{i=0}^{r-1} K_i$ contains a point P of $\prod_{i=2}^{r+1} K_i$. It follows that P is in $\prod_{i=0}^{r+1} K_i$, which by hypothesis is vacuous. Furthermore L_0 contains no point of L_r, $r>2$. For suppose the contrary and that s is the smallest integer greater than two such that L_0 contains a point of L_s. Consider the two continua, L_0+L_s and $L_1+L_2+\cdots+L_{s-1}$. Their common part is the sum of the two mutually exclusive continua, $L_0 \cdot L_1$ and $L_{s-1} \cdot L_s$. Then $\sum_{i=0}^{s} L_i$ is a subcontinuum of M which is the sum of two continua whose common part is disconnected. This is a contradiction of a hypothesis of the theorem. The lemma is therefore true.

LEMMA 2. *If, in a metric space, M is a compact non-degenerate continuum but is not an indecomposable continuum and does not contain a continuum which is the sum of two continua whose common part is disconnected, and if T is*

a reversibly continuous transformation of M into a subset of itself, then T carries some proper subcontinuum of M into itself.

Proof. Suppose T does not carry a proper subcontinuum of M into itself. Then by Lemma 1, there exists a sequence of subcontinua $L_0, L_1, L_2, \cdots$ such that (1) for each positive and negative integer n, L_n is $T^n(L_0)$, (2) L_n and $L_{n\pm1}$ have a point in common, but L_n and L_{n+r}, $|r|>1$, do not have a point in common. Let V_0 be an irreducible continuum from $L_{-1}\cdot L_0$ to $L_0\cdot L_1$ which is a subcontinuum of L_0. Let V_n, for each positive or negative integer, designate $T^n(V_0)$. Then since T is reversibly continuous, it can be easily shown that V_n is irreducible from $L_{n-1}\cdot L_n$ to $L_n\cdot L_{n+1}$. Let P_0 be a point of V_0 not in $L_{-1}\cdot L_0$ or $L_0\cdot L_1$. There is such a point for otherwise L_{-1} would contain a point of L_1. For each positive or negative integer n, let P_n designate $T^n(P_0)$. Let M_0 be an irreducible continuum from P_0 to P_1 which is a subset of $V_0+V_1+L_0\cdot L_1$. Then since T is reversibly continuous, it follows that (1) M_n for each integer n is irreducible from P_n to P_{n+1}, (2) $M_n\cdot M_{n\pm1}$ is not vacuous, but $M_n\cdot M_{n+r}$, $|r|>1$, is vacuous, (3) $(M_i+M_{i-1})\cdot V_i$ is an irreducible continuum from $L_{i-1}\cdot L_i$ to $L_i\cdot L_{i+1}$.

Since M is not indecomposable, M is the sum of two proper subcontinua H and K. One of these continua contains points of infinitely many of the continua $M_0, M_1, M_2, \cdots$. Suppose H contains points of infinitely many of these continua. Let j be the smallest integer greater than or equal to zero such that H contains a point of M_j. Let r be any integer greater than 5 such that H contains a point of M_{j+r}. Let H' designate the common part of H and the continuum $\sum_{i=j}^{j+r}M_i$. But H' is itself a continuum since the common part of two subcontinua of M cannot be disconnected. H' must contain points of $L_{j+1}\cdot L_{j+2}\cdot M_{j+1}$, $L_{j+2}\cdot L_{j+3}\cdot M_{j+2}, \cdots, L_{j+r-2}\cdot L_{j+r-1}\cdot M_{j+r-2}$, since each of these continua separates M_j from M_{j+r} in $\sum_{i=j}^{j+r}M_i$. It follows that H' contains $P_{j+2}, P_{j+3}, \cdots, P_{j+r-2}$, since for each integer s, P_{j+s} lies in a subcontinuum of $\sum_{i=j}^{j+r+1}L_i$ which is irreducible from $L_{j+s-1}\cdot L_{j+s}$ to $L_{j+s}\cdot L_{j+s+1}$. Then H' contains each of the continua $M_{j+2}, M_{j+3}, \cdots, M_{j+r-3}$; since each of these continua is irreducible between two points of $\sum_{i=j+1}^{j+r-2}P_i$. Since r can be taken arbitrarily large, it follows that H' contains M_{j+n} for each integer n greater than 1.

Consider the continuum $C=\overline{\sum_{i=j+2}^{\infty}M_i}$. It is obvious that $T(C)$ is a subset of C. It follows that C is identical with M, for otherwise C contains a proper subcontinuum of M which is carried into itself by T, contrary to hypothesis. But by the argument given above C is a subset of H; therefore H is identical with M. This contradicts the assumption that H is a proper subcontinuum of M. The lemma is therefore true.

3. **A general theorem on fixed points.** We may now prove the general theorem:

THEOREM II. *If M is a compact continuum in a metric space, and if M does not contain an indecomposable continuum and does not contain a continuum which is the sum of two continua whose common part is disconnected, then every reversibly continuous transformation of M into a subset of itself leaves some point of M invariant.*

Proof. By Lemma 2, T carries some proper subcontinuum M_1 of M into itself. Let $M_1, M_2, \cdots, M_\omega, M_{\omega+1}, \cdots$ be a well ordered sequence β of subcontinua of M having the following properties: (1) for each ordinal number θ which has an immediate predecessor, M_θ is a proper subcontinuum of $M_{\theta-1}$ which is carried into itself by T; (2) for each ordinal number ϕ which has no immediate predecessor, M_ϕ is a proper subcontinuum of M_α which is carried into itself by T, where α is any ordinal which is less than ϕ; (3) if M_λ is any non-degenerate continuum which belongs to the sequence β, $M_{\lambda+1}$ is a proper subcontinuum of M_λ belonging to the sequence β. Since M is metric, it is completely separable, and it follows that the sequence β thus defined is countable. Therefore there exists a countable simple subsequence $M_{n1}, M_{n2}, \cdots$ running through β. That is, if M_λ is any element of β, there is an integer i such that M_{ni} is a subcontinuum of M_λ. Since M is compact, the point set $P=\prod_{i=1}^{\infty} M_{ni}$ is a compact continuum and is carried into itself by T, since each of the continua M_{ni} is carried into itself by T. But P cannot be a non-degenerate continuum, for if it is, it has a proper subcontinuum which belongs to the sequence β; and this contradicts the definitions of P and of the sequences β and M_{ni}. It follows that P is a point of M, invariant under T.

4. **Applications to continua in the plane.** We have now the following applications:

THEOREM III. *If M is a compact continuum in the plane which contains no indecomposable continuum, which does not separate the plane, and which contains no domain, then every reversibly continuous transformation of M into a subset of itself leaves some point invariant.*

Proof. It follows from a theorem proved by S. Janiszewski* and also by Miss Mullikan† that a sufficient condition that a compact continuum M separate the plane is that M be the sum of two continua whose common

* S. Janiszewski, *Sur les coupures du plan faites par les continus*, Prace Matematyczno-fizyczne, vol. 26 (1913), pp. 11–63.

† Anna M. Mullikan, *Certain theorems relating to plane connected point sets*, these Transactions, vol. 24 (1922), pp. 144–162.

part is disconnected. Then if M does not separate the plane and contains no domain, no subcontinuum of M separates the plane, and the hypotheses of Theorem II are satisfied. The conclusions of Theorem III then follow.

THEOREM IV. *If D is a bounded simply connected domain in the plane which, together with its boundary, does not separate the plane and whose outer boundary M contains no indecomposable continuum, then every reversibly continuous transformation of $\bar{D}$ into itself leaves some point of $\bar{D}$ invariant.*

Proof. Without loss of generality we may suppose the boundary of D to be identical with its outer boundary, for the compact complementary domain of the outer boundary M of D is itself a simply connected domain which is a subset of $\bar{D}$ and whose boundary is M. Carathéodory* has shown that there exists a conformal and reversibly continuous transformation T_1 of D into the interior I of a given circle J, that there exists a reversibly one-to-one correspondence, also designated by T_1, between the prime ends of D and the points of J, and that this correspondence is characterized as follows: If $P_1, P_2, P_3, \cdots$ is a sequence of points of D converging to a prime end E_P of D (using Carathéodory's definition of convergence), then the sequence of points $T_1(P_1), T_1(P_2), T_1(P_3), \cdots$ converges (in the usual sense) to the point P of J with which E_P is associated by this correspondence; and conversely, if $Q_1, Q_2, Q_3, \cdots$ is a sequence of points of I converging to a point Q of J, then the sequence of points $T_1^{-1}(Q_1), T_1^{-1}(Q_2), T_1^{-1}(Q_3), \cdots$ converges to the prime end E_Q with which Q is associated by the correspondence T_1. Let T be any reversibly continuous transformation of $\bar{D}$ into itself. It can be easily shown that T carries a prime end of D into a prime end of D. Let T_2 be a transformation of the continuum $J+I$ into itself defined as follows: If P is any point of I, let $T_2(P)$ be the point $T_1[T(T_1^{-1}(P))]$. If P is any point of J, let $P_1, P_2, \cdots$ be a sequence of points of I converging to P. Let E_P be the prime end of D associated with P by T_1. Then the sequence $T_1^{-1}(P_1), T_1^{-1}(P_2), T_1^{-1}(P_3), \cdots$ converges to E_P. Since T is reversibly continuous, the sequence $T(T_1^{-1}(P_1)), T(T_1^{-1}(P_2)), \cdots$ converges to the prime end $T(E_P)$ into which E_P is transformed by T. It follows that the sequence of points of I, $T_1[T(T_1^{-1}(P_1))], T_1[T(T_1^{-1}(P_2))], \cdots$ converges to a point of J. That point is defined as $T_2(P)$. Since $T_1[T(T_1^{-1}(P_i))]$ is defined as $T_2(P_i)$, and since T is reversibly continuous on $\bar{D}$, and T_1 is a reversible continuous transformation of D into I, it follows that T_2 is reversibly continuous on I. Also from the definition of T_2 it follows that if $P_1, P_2, P_3, \cdots$ is a sequence of points of I converging to a point P of J, then the sequence $T_2(P_1), T_2(P_2), T_2(P_3), \cdots$

* C. Carathéodory, *Über die Begrenzung einfach zusammenhängender Gebiete*, Mathematische Annalen, vol. 73 (1912), pp. 323–370.

converges to $T_2(P)$, and the sequence $T_2^{-1}(P_1)$, $T_2^{-1}(P_2)$, $T_2^{-1}(P_3)$, $\cdots$ converges to $T_2^{-1}(P)$. Suppose that $Q_1, Q_2, Q_3, \cdots$ is a sequence of points of J converging to a point Q of J. For each positive integer n, let Z_n be a point of I such that each of the distances $d(T_n, Q_n)$ and $d[T_2(Z_n), T_2(Q_n)]$ is less than $1/n$. That such a point Z_n exists is shown as follows: There exists a sequence of points $Z_{n1}, Z_{n2}, Z_{n3}, \cdots$ converging to Q_n. Then by the discussion above, the sequence $T_2(Z_{n1}), T_2(Z_{n2}), T_2(Z_{n3}), \cdots$ converges to $T_2(Q_n)$. There exists an integer i such that $d(Z_{ni}, Q_n) < 1/n$ and $d[T_2(Z_{ni}), T_2(Q_n)] < 1/n$. Then Z_{ni} has the property required of Z_n. We see then that the convergence of the sequence $Q_1, Q_2, \cdots$ to Q implies the convergence of the sequence $Z_1, Z_2, \cdots$ to Q. This in turn implies the convergence of $T_2(Z_1), T_2(Z_2), \cdots$ to $T_2(Q)$ and therefore the convergence of $T_2(Q_1), T_2(Q_2), \cdots$ to $T_2(Q)$. Similarly the sequence of points $T_2^{-1}(Q_1), T_2^{-1}(Q_2), \cdots$ converges to $T_2^{-1}(Q)$, and T_2 is a reversibly continuous transformation of the continuum $J+I$ into itself. By a fundamental theorem on fixed points, T_2 leaves some point of $J+I$ fixed. If T_2 leaves a point of I fixed, then obviously T leaves a point of D fixed. If T_2 leaves a point P of J fixed, then T carries some prime end E_P of D into itself. Let N_P be the continuum in the boundary of D associated with the prime end E_P in the sense that every sequence of points of D converging to E_P has a subsequence which converges in the usual sense to a point of N_P. Then T carries N_P into itself. N. E. Rutt* has shown that a necessary condition that such a continuum as N_P shall be the whole boundary of D is that the boundary of D shall be indecomposable or the sum of two indecomposable subcontinua. Since by hypothesis the boundary of D contains no indecomposable continuum, N_P is a proper subcontinuum of the boundary of D. N_P then is a compact continuum which does not separate the plane and which contains no domain. It satisfies the hypothesis of Theorem III; therefore T leaves some point of N_P fixed. Thus, in any case, T leaves a point of D invariant.

5. **An example.** There exist continua which admit of no continuous transformation into themselves except the identity. We give below an example of a compact acyclic continuous curve in the plane having this property.

Let A_1 be any arc in the plane of length 1. Let A_2 be the sum of A_1 and two arcs of length $1/2^2$ having no point in common with each other or with A_1 except that each has one end point at the midpoint of A_1. For each integer n, let E_{n-1} designate the set of points of A_{n-1} of Menger order greater than two. Let A_n be the sum of A_{n-1} and n arcs of length $1/2^n$ having no point in common with each other or with A_{n-1} except that they all have in common one

* N. E. Rutt, *Prime ends and indecomposability*, Bulletin of the American Mathematical Society, vol. 41 (1935), pp. 265–273.

end point which is the midpoint of a component of $A_{n-1}-E_{n-1}$ whose length is not less than the length of any other component of $A_{n-1}-E_{n-1}$. Let $M=\sum_{r=1}^{\infty}A_r$. Then M is obviously an acyclic continuous curve. Let K be the set of points of M of Menger order greater than two. From the definition of M it follows that no two points of K are of the same Menger order and that K is everywhere dense in M. Let T be any continuous transformation of M into itself. T carries a point of given Menger order into a point of the same Menger order and therefore leaves each point of K fixed. Since K is everywhere dense in M, T leaves each point of M fixed.

It should be noted that M is homeomorphic with a proper subset of itself. E. W. Miller* has given an example of an acyclic continuous curve which is not homeomorphic with any proper subset of itself, but which however does not have the property of the example we have given, since his acyclic continuous curve contains an arc which contains no points of Menger order greater than two. I have not been able to find an example of an acyclic continuous curve in the plane whose only transformation into a subset of itself is the identity.

---

* E. W. Miller, *The Zarankiewicz problem*, Bulletin of the American Mathematical Society, vol. 38 (1932), pp. 831–834.

THE UNIVERSITY OF TEXAS,
AUSTIN, TEXAS

THE TWO CONFORMAL INVARIANTS OF FIFTH ORDER*

BY
EDWARD KASNER

In the problem of classifying curvilinear angles in the plane with respect to the group of conformal transformations, the simplest invariant (other than angular magnitude θ) which occurs is found in the case of a horn angle of first order contact (formula (1)). This is a differential invariant of third order, since it involves third derivatives of the two curves composing the angle. There are no invariants of fourth order (nor of any even order), but there are known to be just two of fifth order, arising in the cases of the horn angle of second order contact and the general right angle, respectively.† The main object of this paper is to determine the two invariants of fifth order explicitly. The results appear in formulas (15) and (23) below.‡ It is shown elsewhere§ how the latter of these expressions can be derived from the former through conformal symmetry. As a corollary we obtain in §3 the invariant of lowest order of a single curve under the inversion group, first found by G. W. Mullins. Generalizations to the case of angles drawn on a curved surface have been given by G. Comenetz.‖

I wish to acknowledge the valuable assistance of Miss A. Vassell and J. De Cicco in calculating the invariants, and of G. Comenetz in connection with §3 and the use of oriented curves.

1. **Horn angle of second order.** We use the term curvilinear angle for the figure formed by an ordered pair of oriented, analytic curve-elements having a common point O, the vertex of the angle.

The symbol γ is employed for the curvature of a curve, γ' for the derivative $d\gamma/ds$ with respect to arc-length s, γ'' for the second derivative $d^2\gamma/ds^2$, and so on. Given a curvilinear angle, the symbols, γ_1, γ_1', γ_1'', $\cdots$ denote the curvature and its successive arc-length derivatives for the first side of the angle, evaluated at the vertex O; and γ_2, γ_2', γ_2'', $\cdots$ are the corresponding

* Presented to the Society, September 10, 1937; received by the editors June 11, 1937. Abstract in Science, vol. 82 (1935), p. 622.

† E. Kasner, *Conformal geometry*, Proceedings of the Fifth International Congress of Mathematicians, Cambridge, 1912, vol. 2, p. 81.

‡ The next simplest invariants, of seventh order, would be encountered in dealing with the horn angle of third order contact and the general 60° angle. The latter is being studied by S. H. Gould and O. P. Doria.

§ See E. Kasner, *Conformal symmetry (Schwarzian reflexion)*, Annals of Mathematics, vol. 38 (1937), pp. 873–879.

‖ Bulletin of the American Mathematical Society, abstract 42-11-399.

quantities for the second side of the angle. The conformal invariants of curvilinear angles given below are all stated in terms of the metric invariants $\gamma_1, \gamma_2, \cdots, \gamma_1''', \gamma_2'''$.

A curvilinear angle is called a horn angle if the two curves coincide in direction at the vertex, and is described as being of kth order contact ($k=1, 2, \cdots$) if the order of contact of the curves at the point of tangency is exactly k. A horn angle of kth order contact is called an H_{k+1}. The conformal invariant of an H_2, as the author has shown, is given by*

$$I_{12} = \frac{\gamma_2' - \gamma_1'}{(\gamma_2 - \gamma_1)^2}. \tag{1}$$

To derive the analogous invariant for an H_3 we proceed as follows.

Suppose that an H_3 lying in the (x, y)-plane is transformed conformally into an H_3 lying in the (X, Y)-plane. We may assume that the coordinates are so chosen in each plane that the vertex of the angle is the origin and the initial tangent direction is that of the positive axis of abscissas. Then in the (x, y)-plane the first side of the angle is given by

$$y = a_2x^2 + a_3x^3 + \cdots \tag{2}$$

and the second side by

$$y = a_2' x^2 + a_3' x^3 + \cdots . \tag{2'}$$

Here $a_2 = a_2'$ and $a_3 \neq a_3'$, because the contact is of second order (exactly). In the (X, Y)-plane the first and second sides of the transformed angle are given respectively by

$$Y = A_2X^2 + A_3X^3 + \cdots \tag{3}$$

and

$$Y = A_2' X^2 + A_3' X^3 + \cdots , \tag{3'}$$

with $A_2 = A_2'$ and $A_3 \neq A_3'$. (All the curves are oriented to the right.)

If we set $z = x+iy$ and $w = X+iY$, the conformal transformation can be represented in some neighborhood of the origin by

$$w = c_1z + c_2z^2 + \cdots , \tag{4}$$

where the coefficients c_n are complex ($c_n = \alpha_n + i\beta_n$), and c_1 is not zero. As the initial direction of the positive axis of abscissas remains fixed, c_1 is necessarily real and positive; $c_1 = \alpha_1 > 0$. To simplify the calculations we shall as-

* See Cambridge paper, and papers by Kasner and Comenetz, Proceedings of the National Academy of Sciences and the American Mathematical Monthly, 1936–1938.

sume at first that $c_1 = 1$, so that instead of (4) we have

$$w = z + c_2 z^2 + c_3 z^3 + \cdots . \tag{4'}$$

It will be easy to allow for the more general transformation (4) later.

Separating (4') into real and imaginary parts, we obtain the equations of the transformation in the real form

$$\begin{aligned} X &= x + \alpha_2(x^2 - y^2) - 2\beta_2 xy + \alpha_3(x^3 - 3xy^2) + \cdots , \\ Y &= y + 2\alpha_2 xy + \beta_2(x^2 - y^2) + \alpha_3(3x^2 y - y^3) + \cdots . \end{aligned} \tag{5}$$

Substituting (5) into (3) and eliminating y by means of (2), we derive an identity in x. If we equate the coefficients of $x^2, \cdots, x^5$, respectively, and solve for $A_2, \cdots, A_5$, after some computation we find the following relations:

$$(6_1)\quad A_2 = a_2 + \beta_2,$$

$$(6_2)\quad A_3 = a_3 + \beta_3 - 2\alpha_2\beta_2,$$

$$(6_3)\quad A_4 = a_4 - \alpha_2 a_3 + 3\beta_2 a_2^2 + (\alpha_3 - \alpha_2^2 + 4\beta_2^2)a_2 + \beta_4 - 2\alpha_3\beta_2 - 3\alpha_2\beta_3 + 5\alpha_2^2\beta_2,$$

$$\begin{aligned}(6_4)\quad A_5 = {} & a_5 - 2\alpha_2 a_4 + 8\beta_2 a_2 a_3 + (\alpha_2^2 + 4\beta_2^2)a_3 + 2\alpha_2 a_2^3 \\ & + (3\beta_3 - 6\alpha_2\beta_2)a_2^2 + (2\alpha_4 - 6\alpha_2\alpha_3 + 4\alpha_2^3 - 24\alpha_2\beta_2^2 + 12\beta_2\beta_3)a_2 \\ & + \beta_5 - 2\beta_2\alpha_4 - 4\alpha_2\beta_4 - 3\alpha_3\beta_3 + 12\alpha_2\beta_2\alpha_3 + 9\alpha_2^2\beta_3 - 14\alpha_2^3\beta_2.\end{aligned}$$

Of course the corresponding relations $(6_1') - (6_4')$ for the second side of the angle are obtained from these by replacing $a_2, \cdots, a_5, A_2, \cdots, A_5$ by $a_2', \cdots, a_5', A_2', \cdots, A_5'$, respectively.

In order to find invariants we must eliminate the coefficients α_n, β_n of the transformation. To eliminate β_5 we subtract (6_4) from $(6_4')$, recalling that $a_2 = a_2'$. The result is

$$\Delta A_5 = \Delta a_5 - 2\alpha_2\Delta a_4 + 8\beta_2 a_2\Delta a_3 + (\alpha_2^2 + 4\beta_2^2)\Delta a_3, \tag{7}$$

where $\Delta A_j \equiv A_j' - A_j$, $\Delta a_j \equiv a_j' - a_j$ $(j = 3, 4, 5)$. Similarly we eliminate β_4 and β_3 by forming the differences

$$\Delta A_4 = \Delta a_4 - \alpha_2\Delta a_3, \tag{8}$$

$$\Delta A_3 = \Delta a_3. \tag{9}$$

Finally α_2 and β_2 are eliminated by substituting from (6_1) and (8) into (7). Rearranging and using (9), we find that

$$\Delta A_3\Delta A_5 - (\Delta A_4)^2 - 4A_2^2(\Delta A_3)^2 = \Delta a_3\Delta a_5 - (\Delta a_4)^2 - 4a_2^2(\Delta a_3)^2. \tag{10}$$

Equations (9) and (10) exhibit two quantities which are invariant under

the special transformation (4′). If we follow (4′) by a magnification

$$w = \alpha_1 z, \qquad \alpha_1 > 0, \tag{11}$$

the resultant is a general conformal transformation of the type (4). It is easy to see that the effect of (11) on a curve (2) is to divide the coefficient a_n by α_1^{n-1}. Consequently (9) and (10) still hold under the general transformation (4) provided we insert the factors $1/\alpha_1^2$ and $1/\alpha_1^6$ in the right members of (9) and (10), respectively.

The coefficients of the curve (2) are expressed in terms of the curvature and its arc-length derivatives at the origin by the standard formulas

$$2a_2 = \gamma_1, \tag{12_1}$$

$$6a_3 = \gamma_1', \tag{12_2}$$

$$24a_4 = \gamma_1'' + 3\gamma_1^3, \tag{12_3}$$

$$120a_5 = \gamma_1''' + 19\gamma_1^2\gamma_1'. \tag{12_4}$$

Using these formulas and the similar ones for the curve (2′), we find that the right members of (9) and (10) can be written, except for numerical factors, as

$$\gamma_2' - \gamma_1' \tag{13}$$

and

$$4(\gamma_2' - \gamma_1')(\gamma_2''' - \gamma_1''') - 5(\gamma_2'' - \gamma_1'')^2 - 4\gamma^2(\gamma_2' - \gamma_1')^2, \tag{14}$$

respectively, where $\gamma \equiv \gamma_1 = \gamma_2$. Now the Jacobian at O of the transformation (4) is α_1^2. Consequently we can state this theorem:

THEOREM I. *The quantities* (13) *and* (14) *are relative invariants of a horn angle of second order contact under conformal transformation. They are of index* 1 *and* 3 *respectively.*

In particular the sign of (13) is invariant, since the Jacobian is positive. The fundamental theorem is as follows:

THEOREM II. *The quantity*

$$\frac{4(\gamma_2' - \gamma_1')(\gamma_2''' - \gamma_1''') - 5(\gamma_2'' - \gamma_1'')^2 - 4\gamma^2(\gamma_2' - \gamma_1')^2}{(\gamma_2' - \gamma_1')^3} \tag{15}$$

and the sign of $\gamma_2' - \gamma_1'$ *are invariants of a horn angle of second order contact under conformal transformation. Conversely, if two such angles agree in the value of* (15) *and the sign of* (13) *they can be made to coincide by a conformal transformation (at least to any finite order of contact).**

* A similar theorem holds for an H_2, the value of (1) and the sign of $\gamma_2 - \gamma_1$ being the corresponding invariants.

The first statement follows from Theorem I. For the proof of the second (converse) statement we refer to the Cambridge paper, where the method of conformal symmetry is applied.

2. **General curvilinear right angle.** A curvilinear right angle is said to be a "general" right angle provided $\gamma_2' + \gamma_1' \neq 0$.*

Let the first side of the angle be given by (2), and the second side by

$$x = b_2y^2 + b_3y^3 + \cdots \tag{16}$$

(oriented upwards). Let the transformed first side be given by (3) and the second side by

$$X = B_2Y^2 + B_3Y^3 + \cdots . \tag{17}$$

Again we begin with a transformation of the special type (4′).

In deducing the equations analogous to $(6_1)-(6_4)$ for the second side of the angle, we can avoid a new calculation if we observe that (4′) can be written as

$$Y + iX = (y + ix) + \frac{1}{i}(\alpha_2 - i\beta_2)(y + ix)^2 + \frac{1}{i^2}(\alpha_3 - i\beta_3)(y + ix)^3 + \cdots .$$

This shows that if we interchange X and Y, x and y, A and B, and a and b, and furthermore replace α_2, β_2, α_3, β_3, α_4, β_4, α_5, β_5, by $-\beta_2$, $-\alpha_2$, $-\alpha_3$, β_3, β_4, α_4, α_5, $-\beta_5$, respectively, then the calculation that was made for $(6_1)-(6_4)$ will turn into the calculation which we now require. Consequently to obtain the new relations we need only make the above interchanges and replacements in equations $(6_1)-(6_4)$. The result is

$$(18_1)\quad B_2 = b_2 - \alpha_2,$$

$$(18_2)\quad B_3 = b_3 + \beta_3 - 2\alpha_2\beta_2,$$

$$(18_3)\quad B_4 = b_4 + \beta_2 b_3 - 3\alpha_2 b_2^2 + (-\alpha_3 + 4\alpha_2^2 - \beta_2^2)b_2 + \alpha_4 - 2\alpha_2\alpha_3 + 3\beta_2\beta_3 - 5\alpha_2\beta_2^2,$$

$$(18_4)\quad B_5 = b_5 + 2\beta_2 b_4 - 8\alpha_2 b_2 b_3 + (4\alpha_2^2 + \beta_2^2)b_3 - 2\beta_2 b_2^3 + (3\beta_3 - 6\alpha_2\beta_2)b_2^2 + (2\beta_4 - 6\beta_2\alpha_3 - 12\alpha_2\beta_3 + 24\alpha_2^2\beta_2 - 4\beta_2^3)b_2 - \beta_5 + 4\beta_2\alpha_4 + 2\alpha_2\beta_4 + 3\alpha_3\beta_3 - 12\alpha_2\beta_2\alpha_3 + 9\beta_2^2\beta_3 - 14\alpha_2\beta_2^3.$$

To eliminate β_5 we add (6_4) and (18_4). Then we subtract $2A_2B_4 + 2B_2A_4$ from the sum to eliminate α_4 and β_4. Next the subtraction of $3(A_2^2 + B_2^2)(A_3 + B_3)/2$

* An equivalent definition is that the curve obtained on reflecting the first side of the angle in the second side (by conformal symmetry or Schwarzian reflection) and then reversing orientation, forms an H_3 with the first side.

removes β_3. Finally we add $-5(A_2^2-B_2^2)(A_3-B_3)/2+2B_2A_2^3+2A_2B_2^3$. In this manner we obtain two expressions which are invariant under the special transformation (4′), namely,

$$b_3 - a_3, \tag{19}$$

$$b_5 + a_5 - 2a_2b_4 - 2b_2a_4 - 4b_2^2b_3 + a_2^2b_3 + b_2^2a_3 - 4a_2^2a_3 + 2a_2b_2^3 + 2b_2a_2^3. \tag{20}$$

Under the general transformation (4), (19) would be multiplied by $1/\alpha_1^2$ and (20) by $1/\alpha_1^4$.

In introducing curvatures for the vertical curve (16) we must be careful to note that the proper formulas are the negatives of $(12_1)-(12_4)$; thus $2b_2=-\gamma_2$, $6b_3=-\gamma_2'$, and so on. Substituting in (19) and (20) we find

$$\gamma_2' + \gamma_1', \tag{21}$$

$$(\gamma_2''' - \gamma_1''') - 5(\gamma_1\gamma_2'' + \gamma_2\gamma_1'') - \gamma_2^2\gamma_2' + 5\gamma_1^2\gamma_2' - 5\gamma_2^2\gamma_1' + \gamma_1^2\gamma_1'. \tag{22}$$

THEOREM III. *The quantities* (21) *and* (22) *are relative conformal invariants of any curvilinear right angle. They are of index* 1 *and* 2, *respectively.*

THEOREM IV. *The quantity*

$$\frac{(\gamma_2''' - \gamma_1''') - 5(\gamma_1\gamma_2'' + \gamma_2\gamma_1'') - \gamma_2^2\gamma_2' + 5\gamma_1^2\gamma_2' - 5\gamma_2^2\gamma_1' + \gamma_1^2\gamma_1'}{(\gamma_2' + \gamma_1')^2} \tag{23}$$

and the sign of $\gamma_2'+\gamma_1'$ *are conformal invariants of a general right angle. Conversely, if two such angles agree in the value of* (23) *and the sign of* (21), *they can be made to coincide by a conformal transformation* (*at least to any finite order of contact*).*

The proof of the converse portion of the theorem is given in the Cambridge paper.

3. **Inversive invarients.** An invariant under the conformal group is of course also an invariant under the *inversion group*, defined by

$$w = \frac{az+b}{cz+d},$$

* If the orientation of the second side is reversed, a general right angle becomes a general 270° angle. The effect of reversing orientation is to change the signs of γ_2, γ_2'', γ_2^{iv}, $\cdots$ but to leave γ_2', γ_2''', $\cdots$ unaltered. It follows that if we replace γ_2 by $-\gamma_2$ and γ_2'' by $-\gamma_2''$, (23) becomes the invariant of a general 270° angle. The same replacements convert (1) and (15) into the invariants of (curvilinear) straight angles of first and second order contact respectively.

All the invariants mentioned in this paper apply also to angles whose sides merely possess fifth derivatives (for (1), third derivatives) at the vertex. This follows from the fact that such an angle can be approximated to fifth (or third) order by an analytic angle, and the fact that a conformal transformation preserves contact. The converses in Theorems II and IV are true only for analytic angles, however.

since the latter is a subgroup of the former.

Consider the horn angle of second order contact formed by a curve and its osculating circle. The value of the invariant (15) in this case is evidently

$$\frac{-4\gamma'\gamma''' + 5\gamma''^2 + 4\gamma^2\gamma'^2}{\gamma'^3}, \tag{24}$$

where $\gamma, \gamma', \cdots$ represent curvature and its arc-length derivatives for the given curve. Now inversion preserves the circle of curvature. Consequently (24) *is an invariant of a single curve under the inversion group.* This was first proved by G. W. Mullins, who showed that (24) is in fact the inversive invariant of lowest order and gave all the higher invariants.*

The question may be raised, for what curves does the value of (1) as formed for the horn angle of first order contact between the curve and its tangent line, remain constant along the curve? (These horn angles at any two points of the curve would then be conformally equivalent, at least to any finite order of contact.) The condition is $-\gamma'/\gamma^2 = \text{const.}$, or $d\rho/ds = \text{const.}$ This is the intrinsic equation of the *logarithmic spirals.*

The analogous question for the horn angle of second order contact between a curve and its osculating circle can be solved by using a theorem of Mullins. In that case the curves are the *logarithmic double spirals*, which may be defined as the transforms by inversion of logarithmic spirals.†

Trihornometry for general horn angles of second order contact will be discussed in later papers. For each order a new type of Finsler space must be introduced. The total geometry is non-archimedian.

---

* *Differential Invariants under the Inversion Group*, Columbia University Dissertation, 1917, p. 21. The I_5 there is identical with (24) above (with $\rho = 1/\gamma$).

† Or as the isogonal trajectories of a pencil of circles through two fixed points, or as the images of loxodromes on a sphere under stereographic projection from an arbitrary point of the sphere (G. Holzmüller, *Ueber die logarithmische Abbildung und die aus ihr entspringenden orthogonalen Curvensysteme*, Zeitschrift für Mathematik und Physik, vol. 16 (1871), p. 269).

COLUMBIA UNIVERSITY,
NEW YORK, N. Y.

THE ABSOLUTE OPTICAL INSTRUMENT*

BY

J. L. SYNGE

1. **Introduction.** An *absolute optical instrument* is a system of transparent media which gives a precise point-image of each object-point lying in a three-dimensional region, the law of light propagation being that of geometrical optics. For a long time only two absolute instruments were known, the plane mirror (or combination of plane mirrors) and the "fish's eye" of Maxwell,† which consists of a single isotropic medium of variable refractive index

$$n = n_0 a^2/(a^2 + r^2), \tag{1.1}$$

where n_0, a are constants, and r is the distance from a fixed point. Maxwell's medium has been generalized by Lenz,‡ and Boegehold and Herzberger§ have given an infinite class of absolute instruments, consisting of homogeneous isotropic media bounded by concentric spheres.

Less attention has been paid to the question of the realization of an absolute instrument by a distribution of transparent media than to the relations between object and image which must hold if the instrument is absolute. In the simplest case, in which the initial and final media are homogeneous and isotropic, it has been shown in various ways by Maxwell, Bruns, Klein, and Liebmann‖ that if A, B are the images of object-points A', B', then

$$n'A'B' = nAB, \tag{1.2}$$

where n', n are the (constant) refractive indices of the initial and final media. This relation may also be written

$$[A'B'] = [AB], \tag{1.3}$$

where the brackets indicate optical lengths.

---

* Presented to the Society, March 27, 1937; received by the editors February 26, 1937 and June 16, 1937.

† J. C. Maxwell, Cambridge and Dublin Mathematical Journal, vol. 8 (1854), p. 188; *Scientific Papers*, vol. 1, Cambridge, 1890, pp. 76–79.

‡ W. Lenz, *Probleme der Moderne Physik*, edited by P. Debye, Leipzig, 1928, pp. 198–207.

§ H. Boegehold and M. Herzberger, Zeitschrift für Angewandte Mathematik und Mechanik, vol. 15 (1935), pp. 157–178.

‖ For references and an historical account by H. Boegehold, see Czapski-Eppenstein, *Grundzüge der Theorie der Optischen Instrumente*, Leipzig, 1924, p. 213 et seq. See also Carathéodory, Sitzungsberichte der Bayerischen Akademie der Wissenschaften, Mathematisch-Physikalische Klasse, 1926, pp. 1–18.

Carathéodory* has extended this result to the case of more general media, namely, to those in which the velocity of propagation of light is the same in opposite directions and this velocity, considered as a function of the direction of a ray, satisfies certain conditions of analyticity.† Further, by introducing the idea of the *field* of the instrument, he made his result applicable to actual instruments.

In the present paper the approach to the problem of the absolute instrument differs from that of Carathéodory. Attention is directed to the surface of components rather than the wave-surface, and we are thereby enabled to see that when the optical character of one of the terminal media is assigned, the optical character of the other is rather closely restricted if the instrument is absolute.

Extension to the case where the space is Riemannian and of N dimensions may easily be made; but since the problem is one of considerable optical interest, it has been thought best not to complicate the presentation by this extension. The problem may also be presented as a problem in Finsler space.

2. **Fundamental theory.** Hamilton's great achievement in geometrical optics was the reconciliation, in the form of a single mathematical treatment, of the emission theory of light and the wave theory,‡ or, in the language of pure mathematics, the calculus of variations and the theory of contact transformations. This comprehensive view places at our disposal for the discussion of any problem in geometrical optics two alternative methods—the ray-method and the wave-method.

The methods employed in the present paper are essentially those of Hamilton, although for brevity an indicial notation will be used, suffixes having the range 1, 2, 3, with the usual summation convention.§ Except for a modification of one of Hamilton's definitions to suit present purposes, there is nothing in the present paper which might not have been given by Hamilton a century ago as an immediate deduction from his theory. The dis-

---

* C. Carathéodory, loc. cit. The part of this paper containing the fundamental theorem is reproduced, with no essential modification, in M. Born, *Optik*, Berlin, 1933, pp. 61–63.

† Professor Carathéodory has informed me by letter that in establishing the result he had exclusively in mind media of the Fresnel type, or media in which similar conditions of analyticity are satisfied. Unfortunately these conditions were not stated in his paper, and the reader might come to the false conclusion that the result follows from the single assumption that the velocity of propagation is the same in opposite directions. That is not so, and it is not difficult to construct artificial examples in illustration. Carathéodory's method is discussed in §9.

‡ Cf. G. Prange, *W. R. Hamiltons Abhandlungen zur Strahlenoptik*, Leipzig, 1933, Anmerkungen, p. 104; *The Mathematical Papers of Sir W. R. Hamilton*, vol. 1, Cambridge, 1931, pp. 277, 497. The latter will be referred to as M. P. H.

§ J. L. Synge, Journal of the Optical Society of America, vol. 27 (1937), pp. 75–82.

cussion of an absolute instrument does not seem to have occurred to him; the idea arose out of Gaussian optics.

The essential parts of the theory will now be developed, care being taken to present them in a form applicable to an absolute instrument. Certain exceptional features are present in such an instrument which may render invalid statements valid in more general cases.

The instrument is composed of any number of media. Each medium has a positive *medium-function* $v(x_r, \alpha_r)$, a function of rectangular cartesian coordinates x_r and of direction cosines α_r; v is homogeneous of degree unity in the direction cosines. It is in general multiple-valued. If the units of space and time are so chosen that the velocity of light in vacuo is unity, then $1/v$ is the *ray velocity* in the medium for a ray having direction cosines α_r at the point x_r. (For an isotropic medium, v is equal to the refractive index.)

By Fermat's principle (accepted as a basic hypothesis) rays satisfy the variational principle $\delta \int v ds = 0$, ds being an element of arc.

The *components of normal slowness* corresponding to a ray through x_r with direction cosines α_r are defined as

$$\sigma_r = \partial v / \partial \alpha_r. \tag{2.1}$$

They were so called by Hamilton* because the vector σ_r stands normal to the wave and its magnitude is equal to the reciprocal of the wave velocity.

The right-hand side of (2.1) being homogeneous of degree zero in the direction cosines, we can eliminate their ratios and obtain the *medium-equation*

$$\Omega(x_r, \sigma_r) = 0. \tag{2.2}$$

If, from an assigned point x_r, we measure off along each straight line with direction cosines α_r a length $1/v$, obtaining a point with relative coordinates α_r/v, the surface so obtained is the *wave-surface* corresponding to x_r. On the other hand if we draw from x_r the totality of vectors σ_r satisfying (2.2), we get the *surface of components* corresponding to x_r. The wave-surface and the surface of components are reciprocal surfaces with respect to the unit sphere with centre x_r.†

* M.P.H., p. 278; M. Herzberger, *Strahlenoptik*, Berlin, 1931, p. 9, calls this vector the "normal vector." "Slowness vector" might be a more descriptive abbreviation for the full title namely, "the vector representing the normal slowness of wave-propagation." These components are of course the optical analogues of the generalized momenta (p_r) of Hamilton's dynamical theory.

† It seems a pity to reject Hamilton's very suggestive terminology, *wave-surface* and *surface of components*, in favor of the names *indicatrix* and *figuratrix*, which carry no intrinsic meaning. Unless the extension to n dimensions appears an important advance, it does not seem historically correct to assign priority in the consideration of these surfaces to Minkowski and Hadamard (cf. C. Carathéodory, *Variationsrechnung*, Leipzig und Berlin, 1935, p. 247). Hamilton had a priority of seventy years, and even he assigned priority to Cauchy.

The *optical length* of any curve C joining points P, Q is

$$[PQ] = \int_C v\,ds, \tag{2.3}$$

the directional arguments of v being the direction cosines of the tangent to C, in the sense P to Q, and ds being a positive element of arc. Giving a weak variation with, in general, displacements of the end points P, Q, we know from the calculus of variations that

$$\delta[PQ] = \left(\frac{\partial v}{\partial \alpha_r}\delta x_r\right)_Q - \left(\frac{\partial v}{\partial \alpha_r}\delta x_r\right)_P - \int_C \left(\frac{d}{ds}\frac{\partial v}{\partial \alpha_r} - \frac{\partial v}{\partial x_r}\right)\delta x_r ds. \tag{2.4}$$

Suppose now that C is a ray; then, by the Euler equations, the integral vanishes and we have, in the notation of (2.1),

$$\delta[PQ] = (\sigma_r \delta x_r)_Q - (\sigma_r \delta x_r)_P. \tag{2.5}$$

The *characteristic function* of an instrument is a function V of the coordinates x_r' of a point A' in the initial medium M' and the coordinates x_r of a point A in the final medium M, such that V is equal to the optical length of a ray joining A' and A. The function V may not be defined for all values of the six variables x_r', x_r corresponding to points A', A in the initial and final media respectively, because there may exist no ray joining A' and A. In general V is multiple-valued, since v is multiple-valued.

We shall use accents to denote quantities pertaining to the initial medium M' of the instrument; quantities pertaining to the final medium will be left unaccented. Different rectangular axes for x_r' and x_r may be used for M' and M respectively. Inspection of the method by which (2.5) has been established shows that the axes to which quantities at P are referred need not be the same as the axes to which quantities at Q are referred. From (2.5) we see at once that if we pass from a ray joining points A', A to an adjacent ray joining points B', B, the variation in V is*

$$\delta V = \sigma_r \delta x_r - \sigma_r' \delta x_r'. \tag{2.6}$$

* This is the fundamental equation of Hamilton's theory, called by him the *equation of the characteristic function* (M.P.H., p. 168). It is of course the optical analogue of the fundamental equation defining contact transformations in dynamics. Hamilton did not frame his theory to take into account the exceptional case presented by an absolute instrument; he assumed that six arbitrary variations $\delta x_r'$, δx_r were permissible in (2.6), so that

$$\partial V/\partial x_r = \sigma_r, \qquad \partial V/\partial x_r' = -\sigma_r'.$$

This cannot be done in the case of an absolute instrument if the points x_r' and x_r correspond to object and image respectively. The argument of the present paper, which makes no such illegitimate assumption, has already been given by Herzberger (op. cit., p. 12) in a different notation.

Following Carathéodory, we shall say that a ray passing through a point A' of M' with direction cosines α_r' *lies in the field of the instrument* if it passes through the instrument to the final medium M. At a given point A' of M' the directions of the rays lying in the field of the instrument will be bounded by a cone of unidirectional straight lines drawn out from A'.

3. **Necessary conditions for an absolute instrument.** Consider the rays emanating in the field of the instrument from an initial point A'. As the medium-functions of the media in the instrument may be multiple-valued, these rays may pass through the final medium in a number of congruences, which we distinguish from one another by calling them congruences of different types.

If all the rays in the final congruence of type T pass through a single point A, we may say that the instrument is *absolute (type T) with respect to the source* A', A being the image (type T) of A'. If the instrument is absolute as above with respect to all initial points and directions in the field of the instrument, we may say that the instrument is *absolute of type T*. If it is absolute for all types, we may say that it is *completely absolute*.

In the present paper we shall be concerned with absolute character for final congruences of a definite type. For simplicity, we shall refer to an instrument possessing this absolute character as *absolute*, without further explicit qualification.

In an absolute instrument there exists a one-to-one correspondence

$$x_r' = x_r'(x_s), \qquad x_r = x_r(x_s'), \tag{3.1}$$

between the points of the initial and final media, A with coordinates x_r being the image of A' with coordinates x_r'.

If, given A', we seek points A such that a ray joins A' to A, we find our choice of A restricted to that part of the final medium traversed by the congruence of rays from A'. Thus, given A' with coordinates x_r', the characteristic $V(x_r', x_r)$ is defined only for a restricted range of values of x_r. This range includes of course the image of A'; the coordinates of that image are, by (3.1), functions of x_r'. When A is chosen at the image of A', there is an infinity of rays joining the two points, but by Fermat's principle their optical lengths are all the same. Hence *in an absolute instrument there exists an absolute characteristic, namely, a function of the three initial coordinates, or of the three final coordinates, whose value is equal to the optical length of a ray joining object to image.* This absolute characteristic will be denoted by F.

Let us now take two adjacent image-points x_r, $x_r+\delta x_r$, and compare the optical lengths of the rays drawn to these points from their respective ob-

jects. The difference between these optical lengths, by (2.6) and (3.1), is given by

$$\begin{aligned} \delta F &= \sigma_r \delta x_r - \sigma_r' \delta x_r' \\ &= \sigma_r \delta x_r - \sigma_s' \frac{\partial x_s'}{\partial x_r} \delta x_r, \end{aligned} \tag{3.2}$$

where σ_r' are the components of any ray (lying in the field) through the object x_r' and σ_r the components of the corresponding ray through the image x_r. Since δx_r are arbitrary, (3.2) gives the three equations

$$\frac{\partial F}{\partial x_r} = \sigma_r - \sigma_s' \frac{\partial x_s'}{\partial x_r}, \tag{3.3}$$

which are the fundamental equations of the present paper.

The implications of (3.3) may be explored in various ways. Thus we may suppose the absolute characteristic F and the transformation (3.1) given, and inquire into the conditions imposed by (3.3) on the optical characters of the initial and final media. Or we may suppose that the initial and final media are of assigned optical characters, given by medium-functions $v'(x_r', \alpha_r')$, $v(x_r, \alpha_r)$ or by surfaces of components

$$\Omega'(x_r', \sigma_r') = 0, \qquad \Omega(x_r, \sigma_r) = 0; \tag{3.4}$$

we may then inquire into the conditions imposed by (3.3) on the absolute characteristic F and the transformation (3.1). The latter type of approach has been the usual one.

4. **The optical character of one terminal medium in an absolute instrument deduced from that of the other.** Let us write (3.3) as

$$\sigma_r = a_{rs}\sigma_s' + b_r, \tag{4.1}$$

where

$$a_{rs} = \frac{\partial x_s'}{\partial x_r}, \qquad b_r = \frac{\partial F}{\partial x_r}. \tag{4.2}$$

Let us consider (4.1) as a transformation expressing final components σ_r in terms of initial components σ_r', the corresponding object and image-points (A' and A) and the absolute characteristic being given. Thus for present purposes a_{rs} and b_r are to be treated as constants; (4.1) is a linear transformation.

Consider now the surface of components S' corresponding to A' (with equation $\Omega'=0$) and the surface of components S corresponding to A (with

equation $\Omega=0$). Consider any ray through A' in the field of the instrument. Let its components be σ_r'. Then the vector drawn from A' with components σ_r' parallel to the axes of x_r' has its end on S'. The corresponding final ray has components given by (4.1). But the vector drawn from A with compo-

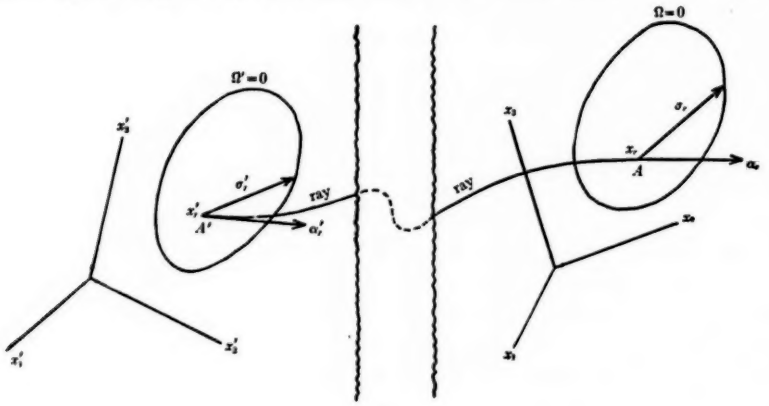


FIG. 1

nents σ_r parallel to the axes of x_r must have its end on S (Fig. 1). Hence we have the following result:

THEOREM I. *In an absolute instrument the portions lying in the field of the instrument of the surfaces of components corresponding to an object-point A' and its image A are transformable into one another by a linear transformation, whose coefficients are in general functions of the coordinates of A'.*

Thus, insofar as the condition (3.3) is concerned, we may choose arbitrarily the following:

(i) the absolute characteristic F,

(ii) the correspondence $x_r=x_r(x_s')$,

(iii) the surfaces of components (that is, the equation $\Omega'(x_r', \sigma_r')=0$) throughout the initial medium.

Then the surfaces of components for the final medium (and hence its medium-function v) can be found to satisfy (3.3). We have in fact by (3.3)

$$\sigma_r' = \sigma_s \frac{\partial x_s}{\partial x_r'} - \frac{\partial F}{\partial x_r'}, \tag{4.3}$$

so that $\Omega'(x_r', \sigma_r')=0$ implies

$$\Omega'\left(x_r'(x_s), \sigma_s \frac{\partial x_s}{\partial x_r'} - \frac{\partial F}{\partial x_r'}\right) = 0; \tag{4.4}$$

these are the required surfaces of components in the final medium, (4.4) being immediately expressible in the form $\Omega(x_r, \sigma_r) = 0$.

In particular we may observe that if there is an absolute instrument in which the surfaces of components in one of the terminal media are ellipsoids (the wave-surfaces being consequently, by the reciprocal relation, ellipsoids also), then the surfaces of components in the other terminal medium are also ellipsoids (and so are the wave-surfaces).

More generally, if in an absolute instrument the surfaces of components throughout one of the terminal media are undegenerate algebraic surfaces of the nth degree, then the surfaces of components throughout the other terminal medium are also algebraic surfaces of the nth degree.

The assumption of algebraic character carries with it an important implication (as would also a suitable assumption regarding analyticity), namely, that a linear transformation which carries the *portion* of one of the surfaces of components corresponding to rays lying in the field of the instrument into a portion of the other surface of components, also carries the *whole* of one surface of components into the whole of the other. It is this fact that enables us to make statements about complete surfaces of components, although our data deal only with the portions corresponding to rays lying in the field of the instrument.

5. **The case where the surfaces of components in the terminal media are general algebraic surfaces.** Let us now suppose that there is an instrument with given terminal media in which the surfaces of components are undegenerate algebraic surfaces of the nth degree. We may write

$$\Omega' \equiv 1 + c'_r \sigma'_r + c'_{rs} \sigma'_r \sigma'_s + \cdots \text{ (to } n+1 \text{ terms)} = 0, \quad (c'_{rs} = c'_{sr}), \tag{5.1}$$

$$\Omega \equiv 1 + c_r \sigma_r + c_{rs} \sigma_r \sigma_s + \cdots \text{ (to } n+1 \text{ terms)} = 0, \quad (c_{rs} = c_{sr}), \tag{5.2}$$

where $c'_r, c'_{rs}, \cdots$ are assigned functions of x'_r, and $c_r, c_{rs}, \cdots$ are assigned functions of x_r.

We wish to investigate the possibility of constructing an absolute instrument with these terminal media, taking into consideration the condition (3.3).

Adopting the notation of (4.1) and substituting in (5.2), we get

$$\Omega \equiv 1 + c_r(a_{rs}\sigma'_s + b_r) + c_{rs}(a_{rt}\sigma'_t + b_r)(a_{su}\sigma'_u + b_s) + \cdots = 0; \tag{5.3}$$

this must be the same as the surface (5.1). Thus we require

$$\begin{aligned} 1 + c_r b_r + c_{rs} b_r b_s + \cdots &= \theta, \\ c_s a_{sr} + 2c_{st} a_{sr} b_t + \cdots &= \theta c'_r, \\ &\cdots\cdots\cdots\cdots, \end{aligned} \tag{5.4}$$

where θ is a factor of proportionality. Here we have a set of equations for five unknown functions of x_r, namely, F, x_r', θ. Omitting from consideration for obvious reasons the case $n=1$, let us pass to the case $n=2$, for which the surfaces are quadrics. There are then ten coefficients in Ω and hence ten equations in (5.4) for five unknowns. If $n>2$, there are of course more equations. *It is therefore in general impossible to construct an absolute instrument with arbitrarily assigned terminal media in which the surfaces of components (or the wave-surfaces) are undegenerate algebraic surfaces of the second degree or of higher degree.*

6. **The case where the surfaces of components in the terminal media are central algebraic surfaces.** By a *central surface* we shall understand a surface which has a centre at the origin; for a surface of components or wave-surface, the origin is the point in the medium to which the surface corresponds.

A linear transformation (4.1) which carries a central surface into a central surface must lack the absolute term. Hence we have this result (of importance in connection with Carathéodory's theorem and implicit in his proof, unless replaced by a condition of analyticity):

THEOREM II. *In an absolute optical instrument in which the surfaces of components (or, equivalently, the wave-surfaces) are undegenerate central algebraic surfaces, the absolute characteristic is a constant. This result also holds for degenerate central surfaces, provided that each of the separate surfaces given by degeneration is central.*

Under these circumstances the transformation of components is simply

$$\sigma_r = \sigma_s' \frac{\partial x_s'}{\partial x_r}. \tag{6.1}$$

7. **The case of isotropic terminal media.** When the terminal media are isotropic, the surfaces of components are spheres with the equations

$$\Omega' \equiv \sigma_r' \sigma_r' - n'^2 = 0, \tag{7.1}$$

$$\Omega \equiv \sigma_r \sigma_r - n^2 = 0, \tag{7.2}$$

where n', n are refractive indices, functions of position. Applying (6.1), we see that if the instrument is absolute, the correspondence of object and image-points must satisfy

$$\frac{\partial x_s'}{\partial x_r} \frac{\partial x_t'}{\partial x_r} = \frac{n^2}{n'^2} \delta_{st}, \tag{7.3}$$

where δ_{st} is the Kronecker delta. These partial differential equations, six in

number, are not in general soluble for the three unknowns x_r'. We deduce from them

$$\frac{\partial x_s'}{\partial x_r} = \frac{n^2}{n'^2}\frac{\partial x_r}{\partial x_s'}, \tag{7.4}$$

and hence

$$\begin{aligned} n^2 ds^2 = n^2 dx_r dx_r &= n^2 \frac{\partial x_r}{\partial x_s'}\frac{\partial x_r}{\partial x_t'} dx_s' dx_t' \\ &= \frac{n'^4}{n^2}\frac{\partial x_s'}{\partial x_r}\frac{\partial x_t'}{\partial x_r} dx_s' dx_t' \\ &= n'^2 dx_s' dx_s' \\ &= n'^2 ds'^2. \end{aligned} \tag{7.5}$$

This establishes (by a method different from, and perhaps more direct than, the method of Carathéodory) the equality of optical lengths for object and image elements in an absolute instrument for which the terminal media are isotropic and in general heterogeneous. A necessary condition for the absolute character of such an instrument is the applicability of the terminal media, considered as Riemannian spaces with metrics nds, $n'ds'$.

8. **The case of Fresnel media.** In a biaxial crystal, or homogeneous Fresnel medium, the surface of components referred to principal axes has the equation*

$$\frac{\sigma_1^2}{1 - c_1^2 q^2} + \frac{\sigma_2^2}{1 - c_2^2 q^2} + \frac{\sigma_3^2}{1 - c_3^2 q^2} = 0, \qquad q^2 \equiv \sigma_r \sigma_r, \tag{8.1}$$

where c_r are constants (the principal velocities). Cleared of fractions, this equation reads

$$q^2 A^0_{mn}\sigma_m\sigma_n - B^0_{mn}\sigma_m\sigma_n + 1 = 0, \qquad q^2 \equiv \sigma_r\sigma_r, \tag{8.2}$$

where

$$\begin{cases} A^0_{11} = c_2^2 c_3^2, & A^0_{22} = c_3^2 c_1^2, & A^0_{33} = c_1^2 c_2^2, \\ B^0_{11} = c_2^2 + c_3^2, & B^0_{22} = c_3^2 + c_1^2, & B^0_{33} = c_1^2 + c_2^2, \end{cases} \tag{8.3}$$

and $A^0_{mn} = B^0_{mn} = 0$ if $m \neq n$.

If we change to an arbitrary set of rectangular axes (not principal) the surface of components has an equation of the form

$$q^2 A_{mn}\sigma_m\sigma_n - B_{mn}\sigma_m\sigma_n + 1 = 0, \qquad q^2 \equiv \sigma_r\sigma_r, \tag{8.4}$$

* Cf. M.P.H., p. 280; Born, op. cit., p. 224.

where $A_{mn}=A_{nm}$, $B_{mn}=B_{nm}$, these quantities being found from A^0_{mn}, B^0_{mn} respectively by application of the rules of tensor transformation. Given an equation (8.4), necessary and sufficient conditions that it should define the surface of components in a homogeneous Fresnel medium are as follows:

(i) The roots of the determinantal equations

$$|A_{mn} - \lambda\delta_{mn}| = 0, \qquad |B_{mn} - \mu\delta_{mn}| = 0, \tag{8.5}$$

are of the form

$$\begin{cases} \lambda_1 = c_2^2 c_3^2, & \lambda_2 = c_3^2 c_1^2, & \lambda_3 = c_1^2 c_2^2, \\ \mu_1 = c_2^2 + c_3^2, & \mu_2 = c_3^2 + c_1^2, & \mu_3 = c_1^2 + c_2^2, \end{cases} \tag{8.6}$$

where the c_r are constants.

(ii) Each of the three perpendicular directions defined by

$$A_{mn}\sigma_n - \lambda_r\sigma_m = 0 \tag{8.7}$$

for fixed r, coincides with the direction defined by

$$B_{mn}\sigma_n - \mu_r\sigma_m = 0 \tag{8.8}$$

for the same value of r.

A heterogeneous Fresnel medium may be defined as a medium in which at each point there exists a set of axes such that the surface of components corresponding to that point has the form (8.1). Equivalently, we may say that, for a single set of rectangular cartesian coordinates throughout the medium, a heterogeneous Fresnel medium is one in which the surface of components at any point has the form (8.4), where A_{mn}, B_{mn} are functions of the coordinates x_r satisfying the conditions (i) and (ii) above, in which c_r are no longer constants but functions of x_r. We shall confine our attention to the undegenerate case in which c_1, c_2, c_3 are distinct.

Let us now consider an absolute instrument in which the terminal media are Fresnel media, in general heterogeneous. Their surfaces of components have the equations

$$q'^2 A'_{mn}\sigma'_m\sigma'_n - B'_{mn}\sigma'_m\sigma'_n + 1 = 0, \qquad q'^2 \equiv \sigma'_r\sigma'_r, \tag{8.9}$$

$$q^2 A_{mn}\sigma_m\sigma_n - B_{mn}\sigma_m\sigma_n + 1 = 0, \qquad q^2 \equiv \sigma_r\sigma_r. \tag{8.10}$$

Since these are undegenerate central algebraic surfaces, the formula for transformation of components is homogeneous of the form (6.1). Hence we must have

$$B_{mn}\sigma_m\sigma_n \equiv B'_{mn}\sigma'_m\sigma'_n, \tag{8.11}$$

$$q^2 A_{mn}\sigma_m\sigma_n \equiv q'^2 A'_{mn}\sigma'_m\sigma'_n. \tag{8.12}$$

Two cases must now be distinguished, according to the way in which we break up (8.12):

Case I:

$$q^2 \equiv \theta^2 q'^2, \qquad A_{mn}\sigma_m\sigma_n \equiv \theta^{-2}A'_{mn}\sigma_m'\sigma_n'; \tag{8.13}$$

Case II:

$$q^2 \equiv \phi^2 A_{mn}\sigma_m'\sigma_n', \qquad A_{mn}\sigma_m\sigma_n = \phi^{-2}q'^2; \tag{8.14}$$

here θ, ϕ are undetermined functions of x_r.

Let us choose a pair of corresponding points (object and image) and choose our axes in the principal directions corresponding. Then (8.11) reads

$$\sum (c_2^2 + c_3^2)\sigma_1^2 \equiv \sum (c_2'^2 + c_3'^2)\sigma_1'^2, \tag{8.15}$$

where $\sum$ indicates a sum of three terms obtained by cyclic permutation of 1, 2, 3. Therefore, renumbering the axes if necessary,

$$\begin{cases} (c_2^2 + c_3^2)\sigma_1^2 \equiv (c_2'^2 + c_3'^2)\sigma_1'^2, \\ (c_3^2 + c_1^2)\sigma_2^2 \equiv (c_3'^2 + c_1'^2)\sigma_2'^2, \\ (c_1^2 + c_2^2)\sigma_3^2 \equiv (c_1'^2 + c_2'^2)\sigma_3'^2. \end{cases} \tag{8.16}$$

Let us now consider separately Case I and Case II.

Case I: Substitution from (8.16) in (8.13) gives

$$\sum \frac{c_2'^2 + c_3'^2}{c_2^2 + c_3^2}\sigma_1'^2 \equiv \theta^2 \sum \sigma_1'^2, \tag{8.17}$$

$$\sum c_2^2 c_3^2 \frac{c_2'^2 + c_3'^2}{c_2^2 + c_3^2}\sigma_1'^2 \equiv \theta^{-2}\sum c_2'^2 c_3'^2 \sigma_1'^2. \tag{8.18}$$

From (8.17) we have

$$\begin{cases} c_2'^2 + c_3'^2 = \theta^2(c_2^2 + c_3^2), \\ c_3'^2 + c_1'^2 = \theta^2(c_3^2 + c_1^2), \\ c_1'^2 + c_2'^2 = \theta^2(c_1^2 + c_2^2), \end{cases} \tag{8.19}$$

from which it follows that

$$c_1' = \theta c_1, \qquad c_2' = \theta c_2, \qquad c_3' = \theta c_3; \tag{8.20}$$

then (8.18) is automatically satisfied.

Case II: Substitution from (8.16) in (8.14) gives

$$\sum \frac{c_2'^2 + c_3'^2}{c_2^2 + c_3^2}\sigma_1'^2 \equiv \phi^2 \sum c_2'^2 c_3'^2 \sigma_1'^2, \tag{8.21}$$

$$\sum c_2^2 c_3^2 \frac{c_2'^2 + c_3'^2}{c_2^2 + c_3^2} \sigma_1'^2 \equiv \phi^{-2} \sum \sigma_1'^2 . \tag{8.22}$$

From (8.21) we have

$$\begin{cases} c_2'^{-2} + c_3'^{-2} = \phi^2(c_2^2 + c_3^2), \\ c_3'^{-2} + c_1'^{-2} = \phi^2(c_3^2 + c_1^2), \\ c_1'^{-2} + c_2'^{-2} = \phi^2(c_1^2 + c_2^2), \end{cases} \tag{8.23}$$

and hence

$$c_1'^{-1} = \phi c_1, \qquad c_2'^{-1} = \phi c_2, \qquad c_3'^{-1} = \phi c_3; \tag{8.24}$$

then (8.22) is automatically satisfied. Hence we have this result:

THEOREM III. *In an absolute instrument in which the terminal media are undegenerate Fresnel media (in general heterogeneous), the principal velocities at a point in one medium are either directly or inversely proportional to the principal velocities at the corresponding point in the other medium.*

9. **Carathéodory's method.** By (3.2), valid for an absolute instrument, we have

$$\delta F = \sigma_r \delta x_r - \sigma_r' \delta x_r' . \tag{9.1}$$

Now it is immediately obvious that if we take adjacent points A', B' on a ray in the initial medium in the field of an absolute instrument, their images A, B lie on the final portion of the ray in question. Taking in (9.1) $\delta x_r'$ to be the displacement from A' to B' and δx_r to be the displacement from A to B, and denoting the direction cosines of initial and final rays as usual by α_r', α_r, we may substitute in (9.1)

$$\delta x_r = \alpha_r ds, \qquad \delta x_r' = \alpha_r' ds', \tag{9.2}$$

where ds, ds' are respectively equal to AB, $A'B'$. Thus we obtain, on division by ds,

$$\frac{\partial F}{\partial x_r} \alpha_r = \sigma_r \alpha_r - \sigma_r' \alpha_r' \frac{ds'}{ds}, \tag{9.3}$$

or, by (2.1) and the homogeneity of v,

$$\frac{\partial F}{\partial x_r} \alpha_r = v(x_r, \alpha_r) - v'(x_r', \alpha_r') \frac{ds'}{ds} . \tag{9.4}$$

But

$$v'(x_r', \alpha_r') \frac{ds'}{ds} = v'\left(x_r', \frac{dx_r'}{ds'}\right) \frac{ds'}{ds} = v'\left(x_r', \frac{dx_r'}{ds}\right) = v'\left(x_r', \frac{\partial x_r'}{\partial x_s} \alpha_s\right), \tag{9.5}$$

and (9.4) thus becomes

$$\frac{\partial F}{\partial x_r}\alpha_r = v(x_r, \alpha_r) - v'\left(x_r', \frac{\partial x_r'}{\partial x_s}\alpha_s\right). \tag{9.6}$$

Except for trivial changes in notation, this is the fundamental equation (14) of Carathéodory's paper (loc. cit.).

Let us now reconstruct Carathéodory's theorem regarding equality of optical lengths, filling in the part of the proof not given explicitly in his paper. We shall assume that the instrument is absolute for rays (in the field) of a certain type, corresponding to a certain one of the values (v') of the multiple-valued medium-function of the initial medium and to a certain one of the values (v) of the multiple-valued medium-function of the final medium. We assume that these single-valued functions v', v are *even analytic* functions of α_r', α_r respectively, throughout three-dimensional spaces in which α_r', α_r are coordinates, except perhaps for certain singular lines through the origins.

Let P be any point in the region of the α_r space which is filled with lines drawn in the directions of rays lying in the field of the instrument, and let Q be the reflection of P in the origin of the α_r space. Let C be an analytic curve joining P and Q. Then if we put

$$\phi(u) = \frac{\partial F}{\partial x_r}\alpha_r - \left\{v(x_r, \alpha_r) - v'\left(x_r', \frac{\partial x_r'}{\partial x_s}\alpha_s\right)\right\},$$

where u is the parameter on C, $\phi(u)$ is an analytic function which vanishes for a range of values of u adjacent to P. Therefore $\phi(u) = 0$ along C and, in particular, at Q. But the first part of $\phi(u)$ is an odd function, and the second part is an even function. Hence, adding and subtracting the values of $\phi(u)$ at P and Q, we get

$$\partial F/\partial x_r = 0, \qquad v(x_r, \alpha_r) = v'\left(x_r', \frac{\partial x_r'}{\partial x_s}\alpha_s\right), \tag{9.7}$$

throughout the field in which P has been taken. Applying the analytic condition again, we see that the second relation holds for all values of α_r, except for the excluded singularities. Hence if $\delta x_r'$, δx_r are object and image elements with lengths $\delta s'$, δs and direction cosines α_r', α_r respectively, we have

$$\begin{aligned} v(x_r, \alpha_r)\delta s &= v'\left(x_r', \frac{\partial x_r'}{\partial x_s}\alpha_s\right)\delta s \\ &= v'\left(x_r', \frac{\partial x_r'}{\partial x_s}\delta x_s\right) \\ &= v'(x_r', \alpha_r')\delta s'. \end{aligned} \tag{9.8}$$

Thus Carathéodory's method establishes the equality of optical lengths for corresponding elements in an absolute instrument. The method by which the fundamental equation (9.1) is obtained in the present paper (indeed, it is deducible in a couple of lines from Hamilton's equation of the characteristic function, properly interpreted) appears more direct than the method given by Carathéodory. The latter part of the argument given above is merely the expansion of Carathéodory's proof, with a full statement of the implied conditions.

The necessary amplification of his original proof will also be found in Carathéodory, *Geometrische Optik*, Berlin, 1937, p. 70, which appeared after the present paper was written. Reference should be made to W. Blaschke, Abhandlungen aus dem Mathematischen Seminar der Hansischen Universität, vol. 11, 1936, pp. 409–412; although not based on the characteristic function, his argument is similar to that of the present paper, for he directs attention to the surface of components and the linear transformation (4.1). I owe this reference to W. C. Graustein.*

* The last paragraph was added in proof, May 20, 1938.

UNIVERSITY OF TORONTO,
TORONTO, CANADA

POLYNOMIAL APPROXIMATIONS FOR ELLIPTIC FUNCTIONS*

BY

E. T. BELL

1. **Equivalences modulo y^{r+1}.** The approximations to be obtained are with respect to k^2, where k is the modulus of the Jacobian elliptic functions. It will be convenient to use the arithmetical concept of congruence, applied to absolutely convergent power series in two independent variables x, y.

All the power series $F(x, y)$, $G(x, y)$, $H(x, y)$, $\cdots$, $P(x, y)$, $\cdots$ considered have a common maximum domain R different from $|xy| = 0$, of absolute convergence; that is, R is the only domain of absolute convergence, different from $|xy| = 0$, of all the series considered, containing R.

The series are given initially in the forms

$$F(x, y) = \sum_{s=0}^{\infty} x^s \left(\sum_{t=0}^{s} f_t(s) y^t \right), \cdots, P(x, y) = \sum_{s=0}^{\infty} x^s \left(\sum_{t=0}^{s} p_t(s) y^t \right), \cdots.$$

they may be rearranged with respect to ascending powers of y; thus, for example,

$$F(x, y) = \sum_{s=0}^{\infty} y^s F^{(s)}(x), \tag{1.1}$$

where

$$F^{(s)}(x) = \sum_{t=0}^{\infty} f_s(s + t) x^{s+t}.$$

Let r be a constant integer $\geqq 0$. Then (1.1) may be written as

$$F(x, y) = F_r(x, y) + y^{r+1} F^{(r+1)}(x, y), \tag{1.2}$$

in which

$$F_r(x, y) = \sum_{s=0}^{r} y^s F^{(s)}(x),$$

$$F^{(r+1)}(x, y) = \sum_{s=0}^{\infty} y^s F^{(r+s+1)}(x).$$

As a function of y, $F_r(x, y)$ is a polynomial of degree $\leqq r$. The degree of the lowest power of y occurring in $F^{(r+1)}(x, y)$ is $\geqq 0$. In analogy with arithmetic, we write (1.2) as

* Presented to the Society, December 28, 1937; received by the editors July 16, 1937.

$$F(x, y) \equiv F_r(x, y) \pmod{y^{r+1}}; \tag{1.3}$$

$F_r(x, y)$ is the *residue* of $F(x, y)$ modulo y^{r+1}. Such residues are the polynomials (in y) with which we shall be concerned. As in arithmetic, we might now pass from congruence as in (1.3) to equality between residue classes. It is more convenient however to proceed as follows: (1.3) is written

$$F(x, y) \sim F_r(x, y), \tag{1.4}$$

which is read, "$F(x, y)$ is *equivalent* to $F_r(x, y)$." The properties of this equivalence follow from (1.2) or (1.3), the latter of which has the usual properties of a congruence relation; the conclusions are restated in the form (1.4). The equivalence in (1.4) has the usual properties of an equivalence relation in algebra. In each of (1.5)–(1.7) the first relation implies the second:

$$\begin{aligned} aF(x, y) + bG(x, y) &= H(x, y), \\ aF_r(x, y) + bG_r(x, y) &\sim H_r(x, y), \end{aligned} \tag{1.5}$$

where a, b are constants;

$$\begin{aligned} F(x, y)G(x, y) &= H(x, y), \\ F_r(x, y)G_r(x, y) &\sim H_r(x, y); \end{aligned} \tag{1.6}$$

$$\begin{aligned} F(x, y)/G(x, y) &= H(x, y), \\ F_r(x, y)/G_r(x, y) &\sim H_r(x, y), \end{aligned} \tag{1.7}$$

provided $G^{(0)}(x)$, in the notation of (1.1), is not zero.

From (1.5)–(1.7) we obtain the equivalence corresponding to any rational relation between $F(x, y), G(x, y), \cdots, P(x, y), \cdots$. Irrational relations also occur, for example $(F(x, y))^\sigma$, written as $F^\sigma(x, y)$, where σ is a positive rational number, may be expanded in the form (1.1), say $F^\sigma(x, y) = P(x, y)$; we then have $F_r^\sigma(x, y) \sim P_r(x, y)$. If $\sigma < 0$, the equivalence holds provided $F^{(0)}(x) \neq 0$.

It will be seen that the residues $F_r(x, y), \cdots$ obtained from elliptic functions can be constructed entirely by finite processes for $r = 0, 1, 2, \cdots$, so that we are operating essentially in the finite domain. The elliptic functions are the infinite series, as $r \to \infty$ through integer values, of the polynomials.

The functions and polynomials introduced in the following section are indispensable for our purpose; they do not seem to have been noticed before.

2. **Functions U, V; polynomials A, B, C, D.** Let a be a non-negative integer. The functions U, V are defined by

$$U_0(x) = \sin x, \qquad U_a(x) = \sum_{s=0}^{\infty} (-1)^s s^a \frac{x^{2s+1}}{(2s+1)!}, \qquad a > 0; \tag{2.1}$$

$$(2.2)\qquad V_0(x) = \cos x, \qquad V_a(x) = \sum_{s=0}^{\infty} (-1)^s s^a \frac{x^{2s}}{(2s)!}, \qquad a > 0.$$

If Δ is the operator of finite differences, and m, n are integers >0, we define $v_n(m)$ by

$$(2.3)\qquad m!v_n(m) = (\Delta^m x^n)_{x=0} = \sum_{i=0}^{m} (-1)^i (m, i)(m-i)^n,$$

where (m, i) is the binomial coefficient $m!/i!(m-i)!$. For $m \leqq n$, $v_n(m)$ is a positive integer; $v_n(m) = 0$ for $m > n$. We shall write

$$s^{[0]} = 1, \; s^{[m]} = s(s-1)\cdots(s-m+1);$$

$$(2.4)\quad u_a(0) = 1; \; u_a(i) = 0, \; i > a; \; u_a(i) = \sum_{j=0}^{a-i} (-1)^j (a, i+j) v_{i+j}(i), \; 0 < i \leqq a.$$

The $u_a(i)$ are further considered in (2.27)–(2.30).

From the binomial expansion of $(-1)^a 2^a s^a = [1-(2s+1)]^a$, (2.1), (2.4), and the known expansion

$$(2.5)\qquad s^a = \sum_{i=1}^{a} v_a(i) s^{[i]}, \qquad a > 0,$$

we obtain

$$(-1)^a 2^a U_a(x) = \sum_{s=0}^{\infty} \frac{(-1)^s x^{2s+1}}{(2s+1)!} \left(1 + \sum_{i=1}^{a} (-1)^i u_a(i)(2s+1)^{[i]}\right).$$

If D_x^i denotes the operator d^i/dx^i, then

$$x^i D_x^i \sin x = \sum_{s=0}^{\infty} (-1)^s (2s+1)^{[i]} \frac{x^{2s+1}}{(2s+1)!};$$

hence, if $a > 0$,

$$(-1)^a 2^a U_a(x) = \sin x + \sum_{i=1}^{a} (-1)^i u_a(i) x^i D_x^i \sin x;$$

therefore, if A, B are defined by

$$(2.6)\qquad A_a(x) = \sum_{i=0}^{\leqq a/2} (-1)^i u_a(2i) x^{2i}, \qquad a > 0,$$

$$(2.7)\qquad B_a(x) = \sum_{i=0}^{\leqq (a-1)/2} (-1)^i u_a(2i+1) x^{2i+1}, \qquad a > 0,$$

we have

$$(2.8)\qquad (-1)^a 2^a U_a(x) = A_a(x)\sin x - B_a(x)\cos x, \qquad a > 0.$$

To (2.6), (2.7) we add the initial values

$$(2.9)\qquad A_0(x) = 1, \qquad B_0(x) = 0,$$

so that (2.8) holds for $a \geqq 0$.

The next equations follow similarly from (2.2) and

$$(2s)^a = \sum_{i=1}^{a} v_a(i)(2s)^{[i]}, \qquad a > 0:$$

$$(2.10)\qquad 2^a V_a(x) = C_a(x)\sin x + D_a(x)\cos x, \qquad a \geqq 0;$$

$$(2.11)\qquad C_0(x) = 0, \qquad C_1(x) = -x, \qquad D_0(x) = 1, \qquad D_1(x) = 0;$$

$$(2.12)\qquad C_a(x) = \sum_{i=1}^{\leqq (a+1)/2} (-1)^i v_a(2i-1)x^{2i-1}, \qquad a > 0;$$

$$(2.13)\qquad D_a(x) = \sum_{i=1}^{\leqq a/2} (-1)^i v_a(2i)x^{2i}, \qquad a > 1.$$

The values of the polynomials for $a = 0, 1, \cdots$ can be calculated directly from (2.7), (2.9), (2.11)–(2.13), but it is easier to obtain them by recurrence. Where the argument x is understood, we shall suppress it, and write $U_a, V_a, \cdots, D_a$ for $U_a(x), \cdots, D_a(x)$. Primes indicate derivatives with respect to x. From (2.1), (2.2) we have at once

$$(2.14)\qquad 2U_{a+1} = xU_a' - U_a, \qquad 2V_{a+1} = xV_a', \qquad a \geqq 0;$$

and by reducing $U_a + xV_a$,

$$(2.15)\qquad 2U_{a+1} + U_a - xV_a = 0, \qquad a \geqq 0.$$

Combining (2.8) with the first of (2.14), and (2.10) with the second, we find that

$$(2.16)\qquad A_{a+1} = A_a - xA_a' - xB_a, \qquad B_{a+1} = B_a - xB_a' + xA_a;$$

$$(2.17)\qquad C_{a+1} = x(C_a' - D_a), \qquad D_{a+1} = x(C_a + D_a').$$

Similarly, from $U_a' = V_a$ we get

$$(2.18)\qquad C_a = (-1)^a(A_a' + B_a), \qquad D_a = (-1)^a(A_a - B_a');$$

and from (2.15),

$$(2.19)\qquad A_{a+1} - A_a + (-1)^a xC_a = 0, \qquad B_{a+1} - B_a - (-1)^a xD_a = 0.$$

The last give

$$(2.20)\qquad \begin{aligned} A_a &= 1 + (-1)^a x(C_{a-1} - C_{a-2} + C_{a-3} - \cdots + (-1)^{a-1}C_0), \\ B &= -(-1)^a x(D_{a-1} - D_{a-2} + D_{a-3} - \cdots + (-1)^{a-1}D_0), \end{aligned}$$

for $a>0$, or for $a\geqq 0$ if we define $C_s=D_s=0$ for $s<0$.

The polynomials satisfy no linear differential equation of order independent of a. Elimination of A or of B from (2.16) gives

$$(2.21)\qquad X_{a+2} - 3X_{a+1} + 2xX'_{a+1} + (x^2+2)X_a - 2xX'_a + x^2X''_a = 0,$$

satisfied by $X=A$, $X=B$; and in the same way

$$(2.22)\qquad Y_{a+2} + Y_{a+1} - 2xY'_{a+1} + x^2Y_a + x^2Y''_a = 0,$$

satisfied by $Y=C$, $Y=D$.

The preceding relations give recurrences for the coefficients u, v in the polynomials. Thus, substituting from (2.12), (2.13) into (2.17) we get the recurrence, equivalent to that for the numbers $\Delta^m 0^n/m!$,

$$(2.23)\qquad v_{a+1}(i) - iv_a(i) - v_a(i-1) = 0;$$

and similarly from (2.16), (2.6), (2.7),

$$(2.24)\qquad u_{a+1}(i) + (i-1)u_a(i) - u_a(i-1) = 0, \qquad a>0.$$

The u are expressed in terms of the v from (2.18),

$$(2.25)\qquad iu_a(i) - u_a(i-1) = (-1)^{a+i}v_a(i-1), \qquad a>0;$$

or from (2.20),

$$(2.26)\qquad u_a(i) = (-1)^a \sum_{j=1}^{\leqq a-i+1} (-1)^i v_{a-j}(i-1), \qquad 1<i\leqq a.$$

For i fixed, (2.24) is in the standard form of the linear difference equation of the first order, with the restriction that $u_0(i)$ is not defined. Solving (2.24) we find

$$u_{a+1}(i) = (1-i)^a\left[\frac{u_2(i)}{1-i} + \sum_{j=2}^{a}\frac{u_j(i-1)}{(1-i)^j}\right], \qquad a>1, \qquad i>1;$$

and hence, since $u_2(2)=1$, $u_2(i)=0$, $i>2$, we have

$$u_{a+1}(2) = (-1)^a\left[-1 + \sum_{j=2}^{a}(-1)^j\right] = \tfrac{1}{2}[1+(-1)^{a+1}], \qquad a>1,$$

which also is seen directly to hold for $a=0$, 1, and

$$(2.27)\qquad u_{a+1}(i) = (1-i)^a \sum_{j=2}^{a}\frac{u_j(i-1)}{(1-i)^j}, \qquad a>1, \qquad i>2.$$

BOSTON UNIVERSITY
COLLEGE OF LIBERAL ARTS
LIBRARY

At once from (2.24),

$$u_a(0) = u_a(1) = 1. \tag{2.28}$$

Taking $i=2$ in (2.27), and noting that $u_1(3)=0$, we get

$$6u_a(3) = -(-1)^a[2^a - (-1)^a - 3], \qquad a > 0; \tag{2.29}$$

whence, with $i=4$ in (2.27),

$$24u_a(4) = (-1)^a[3^a - 2^{a+2} + (-1)^a + 6], \qquad a > 0, \tag{2.30}$$

and so on.

The polynomials are easily calculated recursively from (2.11), (2.16)–(2.19). The first 9 in each set follow. The argument in all is x. These suffice for obtaining approximations to the Jacobian elliptic functions with modulus k up to terms of order k^{16}.

$$A_0 = 1 \qquad B_0 = 0$$
$$A_1 = 1 \qquad B_1 = x$$
$$A_2 = 1 - x^2 \qquad B_2 = x$$
$$A_3 = 1 \qquad B_3 = x - x^3$$
$$A_4 = 1 - x^2 + x^4 \qquad B_4 = x + 2x^3$$
$$A_5 = 1 - 5x^4 \qquad B_5 = x - 5x^3 + x^5$$
$$A_6 = 1 - x^2 + 20x^4 - x^6 \qquad B_6 = x + 10x^3 - 9x^5$$
$$A_7 = 1 - 70x^4 + 14x^6 \qquad B_7 = x - 21x^3 + 56x^5 - x^7$$
$$A_8 = 1 - x^2 + 231x^4 - 126x^6 + x^8 \qquad B_8 = x + 42x^3 - 294x^5 + 20x^7$$

$$C_0 = 0 \qquad D_0 = 1$$
$$C_1 = -x \qquad D_1 = 0$$
$$C_2 = -x \qquad D_2 = -x^2$$
$$C_3 = -x + x^3 \qquad D_3 = -3x^2$$
$$C_4 = -x + 6x^3 \qquad D_4 = -7x^2 + x^4$$
$$C_5 = -x + 25x^3 - x^5 \qquad D_5 = -15x^2 + 10x^4$$
$$C_6 = -x + 90x^3 - 15x^5 \qquad D_6 = -31x^2 + 65x^4 - x^6$$
$$C_7 = -x + 301x^3 - 140x^5 + x^7 \qquad D_7 = -63x^2 + 350x^4 - 21x^6$$
$$C_8 = -x + 966x^3 - 1050x^5 + 28x^7 \qquad D_8 = -127x^2 + 1701x^4 - 266x^6 + x^8$$

Although they will not be required here, it may be mentioned that generalizations of $U_a(x)$, $V_a(x)$ to non-integral values of a, by means of contour integrals, have been investigated by Professor H. Bateman. These lead to expansions of the functions (for any a) in terms of Bessel functions; such expansions offer a point of departure for approximating the values of the polynomials in $y(=k^2)$ next obtained as approximations in y to $\mathrm{sn}(x, k)$, $\mathrm{cn}(x, k)$,

$\mathrm{dn}(x, k)$, for large values of the x occurring in the coefficients of the polynomials.

3. **The polynomials** $\mathrm{sn}_r x$. The modulus of sn x being k, we write $\mathrm{sn}(x, k) = \mathrm{sn}(x|k^2)$, in Milne-Thomson's notation,* as it is usually k^2, not k, that is given in applications, and only even powers of k appear in the power series expansion of $\mathrm{sn}(x, k)$. Finally we write $k^2 = y$, $1 - y = y_1 = k'^2$, and consider $\mathrm{sn}(x, k) = \mathrm{sn}(x|y) = \mathrm{sn}\, x$ as a function of x, y. Similarly for cn x, dn x. For the meaning of $\mathrm{sn}_r x$, $\cdots$, see §1. We shall use the forms of the power series obtained in a previous paper.† For sn x we found

$$\mathrm{sn}\, x = \sum_{s=0}^{\infty}\Big[\sum_{r=0}^{s} p_r(s) y^r\Big]\frac{(-1)^s x^{2s+1}}{(2s+1)!},$$

$$p_0(s) = 1,\quad 2^4 p_1(s) = 3^{2s+1} - 8s - 3,$$

$$2^8 p_2(s) = 5^{2s+1} - 4(2s-1)3^{2s+1} + 32s^2 - 32s - 17,$$

$$2^{12} p_3(s) = 7^{2s+1} - 4(2s-3)5^{2s+1} + (32s^2 - 88s + 30)3^{2s+1} - \tfrac{1}{3}(256s^3 - 1056s^2 + 752s + 471),$$

the general $p_j(s)$ being of the form

$$(3.1)\qquad 2^{4j} p_j(s) = (2j+1)^{2s+1} + P_{j1}(s)(2j-1)^{2s+1} + \cdots + P_{jj}(s)1^{2s+1},$$

in which $P_{jt}(s)$ is a polynomial in s of degree t with rational coefficients. Moreover it was shown (loc. cit., p. 846, (8)) that the $p_j(s)$ can be calculated by linear recurrence, and in a subsequent paper‡ numerous linear recurrences were given for the calculation of the coefficients appearing in the recurrences for the $p_j(s)$. We may therefore consider the $p_j(s)$ for $j = 0, 1, 2, \cdots$ known, as it is straight-forward elementary algebra to obtain a particular $p_j(s)$ by the means indicated.

For the meaning of $\mathrm{sn}_r x$ see (1.2). The sx occurring in the sin sx, cos sx appearing in $\mathrm{sn}_r x$ expresses a number of radians. We have

$$(3.2)\qquad \mathrm{sn}_0\, x = \sum_{s=0}^{\infty} p_0(s)\frac{(-1)^s x^{2s+1}}{(2s+1)!} = \sin x;$$

$$\mathrm{sn}_1\, x = \sum_{s=0}^{\infty}[p_0(s) + p_1(s)y]\frac{(-1)^s x^{2s+1}}{(2s+1)!}.$$

Reducing the coefficient of y in the last we get (see §2)

* L. M. Milne-Thomson, *Die elliptischen Funktionen von Jacobi* (5-figure tables), Berlin, 1931.

† These Transactions, vol. 36 (1934), pp. 841–852, in which note the following misprints: p. 842 (1), for x^2 read x^{2s}; p. 843, last line, for 2^{2s} read 3^{2s}; p. 844, in the expression for $q_3(s)$, for -297 read $+297$; p. 844 (7), all exponents on the right should be $2s$, not s.

‡ American Journal of Mathematics, vol. 48 (1936), pp. 759–768.

$$2^4\sum_{s=0}^{\infty} p_1(s)\frac{(-1)^s x^{2s+1}}{(2s+1)!} = \sum_{s=0}^{\infty}(3^{2s+1}-8s-3)\frac{(-1)^s x^{2s+1}}{(2s+1)!}$$
$$= \sin 3x - 8U_1(x) - 3\sin x$$
$$= \sin 3x - 3\sin x + 4(A_1 \sin x - B_1 \cos x)$$
$$= \sin 3x + \sin x - 4x\cos x;$$

$$\text{(3.21)} \qquad \mathrm{sn}_1 x = \sin x + 2^{-4}y(\sin 3x + \sin x - 4x\cos x).$$

Proceeding to $\mathrm{sn}_2 x$ we have the new term in y^2. The coefficient of $2^{-8}y^2$ is

$$\sum_{s=0}^{\infty}\frac{(-1)^s x^{2s+1}}{(2s+1)!}[5^{2s+1} - (8s-4)3^{2s+1} + 32s^2 - 32s - 17],$$
$$= \sin 5x + 4\sin 3x - 17\sin x - 8U_1(3x) + 32U_2(x) - 32U_1(x).$$

The new detail $U_1(3x)$ is typical of the like in subsequent calculations. By (2.8),

$$-8U_1(3x) = 4[A_1(3x)\sin 3x - B_1(3x)\cos 3x],$$
$$= 4(\sin 3x - 3x\cos 3x).$$

The remaining terms are evaluated as before, and we get

$$\text{(3.3)} \qquad \begin{aligned}\mathrm{sn}_2 x &= \sin x + 2^{-4}y(\sin 3x + \sin x - 4x\cos x)\\ &+ 2^{-8}y^2[\sin 5x + 8\sin 3x + (7-8x^2)\sin x - 12\cos 3x - 24x\cos x].\end{aligned}$$

These are enough to show the process.

By (1.2) and the cited expansion of sn x we have, for $r>0$,

$$\mathrm{sn}_r x = \mathrm{sn}_{r-1} x + y^r\sum_{s=0}^{\infty} p_r(s)(-1)^s\frac{x^{2s+1}}{(2s+1)!},$$

and hence by (3.1) the coefficient of $2^{-4r}y^r$ in $\mathrm{sn}_r x$ is

$$\text{(3.4)} \qquad \sum_{s=0}^{\infty}\frac{(-1)^s x^{2s+1}}{(2s+1)!}\left[(2r+1)^{2s+1} + \sum_{t=1}^{r} P_{rt}(s)(2r+1-2t)^{2s+1}\right],$$

in which

$$\text{(3.5)} \qquad P_{rt}(s) = p_{t0}(r)s^t + p_{t1}(r)s^{t-1} + \cdots + p_{tt}(r), \qquad t>0,$$

the $p_{tj}(r)$ being rational numbers; by convention we take $p_{00}(r)=1$. From this we can determine the general form of $\mathrm{sn}_r x$. The result of substituting from (3.5) into (3.4) and reducing by (2.1) is

$$\text{(3.6)} \qquad \sin(2r+1)x + \sum_{t=1}^{r}\sum_{j=0}^{t} p_{tj}(r)U_{t-j}((2r+1-2t)x).$$

To express this in the same form as (3.3) we apply (2.8), noting that $p_{00}(r)=1$, and defining the polynomials α, β by

$$\alpha_t(r, x) = \sum_{j=0}^{t} (-2)^{j-t} p_{tj}(r) A_{t-j}((2r+1-2t)x),$$

$$\beta_t(r, x) = \sum_{j=0}^{t} (-2)^{j-t} p_{tj}(r) B_{t-j}((2r+1-2t)x), \qquad t \geqq 0.$$

Then the whole expression in (3.6) becomes

$$\sum_{t=0}^{r} [\alpha_t(r, x) \sin (2r+1-2t)x - \beta_t(r, x) \cos (2r+1-2t)x].$$

Hence finally, for $r>0$,

$$\operatorname{sn}_r x = \sin x + \sum_{i=1}^{r} 2^{-4i} y^i \sum_{t=0}^{i} [\alpha_t(i, x) \sin (2r+1-2t)x - \beta_t(i, x) \cos (2i+1-2t)x]. \tag{3.7}$$

For a given r, and t an integer >0, $\operatorname{sn}_r^t x$ can be calculated directly from $\operatorname{sn}_r x$. Powers of sines and cosines are expressed as sums of sines or cosines of multiple angles. For $t=r=2$ the result is, by (3.3),

$$\begin{aligned}\operatorname{sn}_2^2 x = {} & 2^{-1}(1-\cos 2x) + 2^{-4}y(1-\cos 4x - 4x \sin 2x) \\ & + 2^{-9}y^2[16 + (3+32x^2)\cos 2x - 16\cos 4x - 3\cos 6x \\ & - 32x \sin 4x - 40x \sin 2x].\end{aligned} \tag{3.8}$$

4. **The polynomials** $\operatorname{cn}_r x$. The preliminary expansion is

$$\operatorname{cn} x = 1 + \sum_{s=1}^{\infty} \left[\sum_{r=0}^{s-1} q_r(s) y^r \right] \frac{(-1)^s x^{2s}}{(2s)!},$$

$$\begin{aligned} q_0(s) &= 1, \qquad 2^4 q_1(s) = 3^{2s} - 8s - 1, \\ 2^8 q_2(s) &= 5^{2s} - 8(s-1)3^{2s} + 32s^2 - 48s - 9, \\ 2^{12} q_3(s) &= 7^{2s} - 8(s-2)5^{2s} + 2(16s^2 - 60s + 41)3^{2s} \\ &\quad - \tfrac{1}{3}(256s^3 - 1248s^2 + 1280s + 297). \end{aligned}$$

Corresponding to (3.1), (3.5) we have

$$2^{4r} q_r(s) = (2r+1)^{2s} + Q_{r1}(s)(2r-1)^{2s} + \cdots + Q_{rr}(s)1^{2s}, \qquad r>0,$$

$$Q_{rt}(s) = q_{t0}(r)s^t + q_{t1}(r)s^{t-1} + \cdots + q_{tt}(r),$$

the $q_{tj}(r)$ being rational numbers, and $q_{00}(r)=1$ by convention.

Proceeding as in §3 we find

$$\begin{aligned}\mathrm{cn}_2\, x &= \cos x + 2^{-4}y(\cos 3x - \cos x + 4x \sin x)\\ &\quad + 2^{-8}y^2[\cos 5x + 8\cos 3x - (9 + 8x^2)\cos x\\ &\quad + 12x \sin 3x + 16x \sin x].\end{aligned} \tag{4.1}$$

This may also be obtained (see §1) from $\mathrm{cn}_r^2\, x + \mathrm{sn}_r^2\, x \sim 1$; whence

$$\mathrm{cn}_2\, x = (1 - \mathrm{sn}_2^2\, x)^{1/2} = 1 - \tfrac{1}{2}\,\mathrm{sn}_2^2\, x - \tfrac{1}{8}\,\mathrm{sn}_2^4\, x - \cdots.$$

The value of $\mathrm{cn}_2^2\, x$ can be determined directly from (4.1), as for $\mathrm{sn}_2^2\, x$ in §3. The result is (as it should be) $1 - \mathrm{sn}_2^2\, x$, where $\mathrm{sn}_2^2\, x$ is given in (3.8). To indicate the general form corresponding to (3.6), we define the polynomials γ, δ:

$$\gamma_t(r, x) = \sum_{j=0}^{t} 2^{j-t} q_{tj}(x) C_{t-j}((2r + 1 - 2t)x),$$

$$\delta_t(r, x) = \sum_{j=0}^{t} 2^{j-t} q_{tj}(x) D_{t-j}((2r + 1 - 2t)x), \qquad t \geqq 0,$$

and find, for $r > 0$,

$$\begin{aligned}\mathrm{cn}_r\, x = \cos x + \sum_{i=1}^{r} 2^{-4i} y^i \sum_{t=0}^{i} [&\gamma_t(i, x) \sin (2i + 1 - 2t)x\\ &+ \delta_t(i, x) \cos (2i + 1 - 2t)x].\end{aligned} \tag{4.2}$$

In getting (4.2) we have used the second of the following formulas, which follows from the first,

$$\mathrm{cn}_r^2\, x + \mathrm{sn}_r^2\, x \sim 1, \quad \sum_{t=0}^{r} q_{tt}(r) = 0, \qquad r > 0.$$

5. **The polynomials** $\mathrm{dn}_r\, x$. The preliminary expansion is

$$\mathrm{dn}\, x = 1 + \sum_{s=1}^{\infty}\left[\sum_{r=0}^{s-1} h_r(s) y^{r+1}\right]\frac{(-1)^s x^{2s}}{(2s)!},$$

$$2^2 h_0(s) = 2^{2s}, \qquad 2^6 h_1(s) = 2^{2s}(2^{2s} - 8s + 4),$$

$$2^{10} h_2(s) = 2^{2s}[3^{2s} - 4(2s - 3)2^{2s} + 32s^2 - 88s + 31],$$

the general coefficient being of the form

$$h_r(s) = 2^{2s-4r-2}[(r + 1)^{2s} + H_{r1}(s) r^{2s} + \cdots + H_{rr}(s) 1^{2s}],$$

$$H_{rt}(s) = h_{t0}(r) s^t + h_{t1}(r) s^{t-1} + \cdots + h_{tt}(r);$$

the $h_{tj}(r)$ are rational numbers, and by convention $h_{00}(r) = 1$. As before we get

$$\begin{aligned}\mathrm{dn}_2\, x &= 1 - 2^{-2}y(1 - \cos 2x)\\ &\quad + 2^{-6}y^2(-5 + 4\cos 2x + \cos 4x + 8x \sin 2x).\end{aligned} \tag{5.1}$$

The identity $\mathrm{dn}^2 x = 1 - y\,\mathrm{sn}^2 x$ gives

$$\mathrm{dn}_r^2 x \sim 1 - y\,\mathrm{sn}_r^2 x\,;$$

hence in particular, by (3.8),

$$\mathrm{dn}_2^2 x = 1 - 2^{-1}y(1 - \cos 2x) + 2^{-4}y^2(-1 + \cos 4x + 4x \sin 2x). \tag{5.2}$$

In the same way as before we find the general form,

$$\mathrm{dn}_0 x = 1, \qquad \mathrm{dn}_1 x = 1 + 2^{-2}y(-1 + \cos 2x), \tag{5.3}$$

with ϵ, ζ defined by

$$\epsilon_t(r, x) = \sum_{j=0}^{t} 2^{j-t} h_{tj}(r) C_{t-j}(2(r + 1 - t)x),$$

$$\zeta_t(r, x) = \sum_{j=0}^{t} 2^{j-t} h_{tj}(r) D_{t-j}(2(r + 1 - t)x), \qquad t \geqq 0\,;$$

and, for $r > 0$,

$$\begin{aligned}\mathrm{dn}_{r+1} x = {} & 1 + 2^{-2}y(-1 + \cos 2x) \\ & + \sum_{i=1}^{r} 2^{-4i-2} y^{i+1} \sum_{t=0}^{i} [-h_{tt}(i) + \epsilon_t(i, x) \sin (2(i + 1 - t)x) \\ & \qquad + \zeta_t(i, x) \cos (2(i + 1 - t)x)].\end{aligned} \tag{5.4}$$

The approximations in this and preceding sections give approximations to the standard elliptic integrals. Thus from

$$E(x) = \int_0^x \mathrm{dn}^2 x\, dx$$

and (5.2) we get

$$E_2(x) = x - 2^{-1}y(x - \sin x) - 2^{-6}y^2(4x - 4 \sin 2x - \sin 4x + 8x \cos 2x)$$

for the elliptic integral of the second kind.

California Institute of Technology,
Pasadena, Calif.

PLANE PEANIAN CONTINUA WITH UNIQUE MAPS ON THE SPHERE AND IN THE PLANE*

BY
V. W. ADKISSON

INTRODUCTION

The plane peanian continuum† M is said to have a unique map on a spherical surface (or a plane) if, and only if, for any topological image M' of M on a sphere (or plane) S' and any topological image M'' of M on a sphere (or plane) S'', every homeomorphism of M' into M'' can be extended to a homeomorphism of S' into S''.

It is the purpose of this paper to characterize the plane peanian continua that have unique maps on the sphere or in the plane.‡

DEFINITIONS. The simple closed curve J of a cyclicly connected continuum C is called a *bounding circuit* of C provided that for any two maximal connected components H and K of $C-J$ the sets $\overline{H}\cdot J$ and $\overline{K}\cdot J$ lie respectively on two distinct arcs AXB and AYB of J.§

A *split circuit* is a bounding circuit J such that $C-J$ contains at least two components.

If C is cyclicly connected, then C is *triply connected* if, and only if, it is impossible to express C as the sum of two closed connected sets A and B such that neither A nor B is an arc and the set $A\cdot B$ consists of two distinct points of C.‖

Theorems IV and V state the principal results of this paper.

THEOREM IV. *The plane peanian continuum M has a unique map on the sphere if, and only if, one of the following conditions holds*:

* Presented to the Society, September 9, 1937; received by the editors July 22, 1937.

† A plane peanian continuum is a peanian continuum (continuous curve) that has a map (topological image) in the plane or on the sphere. For a characterization of these continua see W. S. Claytor, *Topological immersion of peanian continua in a spherical surface*, Annals of Mathematics, vol. 35 (1934), pp. 809–835. See also Claytor, *Peanian continua not imbeddable in a spherical surface*, ibid., vol. 38 (1937), pp. 631–646.

‡ This problem was suggested by J. R. Kline. The author also wishes to express his appreciation of suggestions by Saunders MacLane which have led to improvements in this paper.

§ For cyclicly connected continua *bounding circuit* is equivalent to *boundary curve* as defined by Claytor, loc. cit., first paper, p. 809. A simple closed curve J of M is called a boundary curve of M provided that there do not exist in $M-J$ distinct components H and K such that (1) a point pair of $\overline{H}\cdot J$ separates a point pair of $\overline{K}\cdot J$ on J, or (2) $\overline{H}\cdot J=\overline{K}\cdot J=$three distinct points. Both definitions will be found useful in later proofs.

‖ Cf. definition of triply connected graphs, H. Whitney, *Congruent graphs and the connectivity of graphs*, American Journal of Mathematics, vol. 54 (1932), p. 158.

(1) *M is acyclic and consists of either a simple arc or a triod.**

(2) *M contains one cyclic element C† which is a maximal triply connected cyclic curve of M, and $M-C$ consists of at most a countable number of arcs, $a_1, a_2, a_3, \cdots$, such that $\bar{a}_i \cdot \bar{a}_j = 0$, $(i \neq j)$, and each $\bar{a}_i \cdot C$ is a single point which lies on only one bounding circuit of C,‡ provided that if C is a simple closed curve, then $M-C$ is at most a simple arc.*

THEOREM V. *The plane bounded peanian continuum M has a unique map in the plane if, and only if, M is one of the following curves:*

(1) *a simple arc,*

(2) *a triod,*

(3) *a simple closed curve,*

(4) *a curve M such that M contains a closed 2-cell C and $M-C$ consists of at most a countable number of arcs $a_1, a_2, a_3, \cdots$, such that $\bar{a}_i \cdot \bar{a}_j = 0$, $(i \neq j)$, and each $\bar{a}_i \cdot C$ is a single point which lies on the only bounding circuit of C.*§

PRELIMINARY THEOREMS

THEOREM I. *The plane cyclic (cyclicly connected) peanian continuum C is triply connected if, and only if, C contains no split circuit.*

THEOREM II. *The plane cyclic peanian continuum C has a unique map on the sphere if, and only if, C is triply connected.*

COROLLARY 1. *If J is a bounding circuit of C which is not a split circuit, then in every map of C on the sphere (or plane) the image of J is the boundary of a complementary domain of the map of C.*

COROLLARY 2. *If J is a split circuit of C there is some map of C on the sphere (or plane) in which the image of J is the boundary of a complementary domain of the map of C.*‖

DEFINITION. The point p of M is a *split-point* of M, if and only if, M can be expressed as the sum of two closed connected sets A and B such that $A \cdot B = p$, and such that if A or B is an arc, then p is not an end point of that arc.

THEOREM III. *If the plane peanian continuum M has a unique map on the sphere or in the plane, then M does not contain a split-point.*

---

* Three arcs PA, PB, PC with P, and only P, common to any two.

† G. T. Whyburn, *Concerning the structure of a continuous curve*, American Journal of Mathematics, vol. 50 (1928), p. 168.

‡ Each point $\bar{a}_i \cdot C$ is a non-local cut-point of C. See Theorem VI and the alternative statement of Theorem IV at the end of this paper.

§ See Claytor. loc. cit., p. 810, Corollary (C).

‖ This is a generalization of Proposition K, Claytor, loc. cit., first paper, p. 828.

Proofs of Theorems I–V

Proof of Theorem I. Let C be a map of the given curve on a plane S, and suppose C not triply connected. There exist two points p and q of C such that $C=A+B$ and $A\cdot B=p+q$, where A and B are closed connected sets neither of which is an arc. Let r be a point of $A-p-q$ and s a point of $B-p-q$. Then on any arc rxs lying in $S-p-q$ there is a last point of A from r to s and a first point of B. Let these points be r and s respectively. The open arc $\langle rxs\rangle$ lies in a domain R complementary to C with boundary $J\supset r+s$. Hence this circuit J must pass through p and q. But since neither A nor B is an arc there must exist at least two components of $C-J$; one in $A-A\cdot J$ and one in $B-B\cdot J$. Let N_1 and N_2 be any two components of $C-J$. Then two points of $\overline{N}_1\cdot J$ cannot separate two points of $\overline{N}_2\cdot J$ on J. For N_1 and N_2 are connected sets lying in $S-\overline{R}$ and cannot intersect. If $\overline{N}_1\cdot J=\overline{N}_2\cdot J=$three points, then N_1 and N_2 must intersect; this is impossible. Therefore, since $C-J$ contains at least two components, J is a split circuit.

Now assume that C contains a split circuit J. There is a component N_1 of $C-J$ with limit points all on the arc gxh of J and a component N_2 with limit points all on the arc $gyh=J-\langle gxh\rangle$. The arc gxh may be chosen so that g and h are points of $\overline{N}_1\cdot J$. Now every component of $C-J$ must have its limit points in either gxh or gyh. For suppose some component N of $C-J$ has a limit point d in $\langle gxh\rangle$ and a limit point e in $\langle gyh\rangle$. Every arc of J containing d and e will necessarily contain either g or h, limit points of N_1 which separate d and e on J. But this contradicts the fact that J is a split circuit.

Now let $A=gxh+N_1$ plus all components of $C-J$ (different from N_2) with limit points on gxh, and let $B=gyh+N_2$ plus all components of $C-J$ not included in A. Then $C=A+B$, where A and B are closed connected subsets of C; neither A nor B is a simple arc; and $A\cdot B=g+h$. Therefore C is not triply connected. This completes the proof of Theorem I.

Proof of Theorem II. Suppose C is triply connected. Then every bounding circuit of C has an image which is a c.d.b.* In every map of C. For let C' be any map of C on a sphere S', and J' any bounding circuit of C'. If C' does not consist of a simple closed curve, then there is one, and only one, component of $C'-J'$ (Theorem I), and this component must lie entirely in one of the regions of S' bounded by J'. The other region of S' bounded by J' must be a complementary domain of C'. Obviously every c.d.b. of C' is also a bounding circuit of C.

Therefore if C' is a map of C on a sphere S' and C'' a map of C on S'', every homeomorphism of C' into C'' must preserve complementary domain

* "Complementary domain boundary" will be abbreviated "c.d.b."

boundaries and is extendable to S' and S''.* Hence C has a unique map on the sphere.

Suppose C is not triply connected. Then C must contain a split circuit J (Theorem I). We shall show first that there is a map of C in which the map of J is a c.d.b. Let C be any map of C on a sphere S and suppose J is not a c.d.b. of C. Let R_1 and R_2 be the two regions of S bounded by J. Since J is not a c.d.b. of C there must exist a component N_1 of $C-J$ in R_1 and a component N_2 in R_2. Any two points p and q of $\overline{N}_1 \cdot J$ lie on some c.d.b. of C within $\overline{R}_2$. For if every arc from p to q lying in $\overline{R}_2$ contains a point of C, there would then exist a connected subset of C lying in R_2 with limit points on J that separate p and q.† But such a connected subset would belong to a component of $C-J$, and J would not be a split circuit. Hence p and q lie on some c.d.b. within $\overline{R}_2$. Furthermore any three points p, q, r of $\overline{N}_1 \cdot J$ lie on the same c.d.b. of C within $\overline{R}_2$. To show this suppose p, q, r do not lie on the same c.d.b. within $\overline{R}_2$. Then there are three complementary domains of C in $\overline{R}_2$ each containing a pair of p, q, r on its boundary J_i, $(i=1, 2, 3)$. Let l, m, n be three arcs, one from each of the boundaries J_i such that $l+m+n \supset p+q+r$ and bounds a region R_3 which is a subset of R_2 containing no points of $J_1+J_2+J_3$. This is possible since no two arcs of l, m, n can have a common point. For if two arcs, l, m do have a common point (other than an end point) there would exist a component $N_3 \supset (l+m)$ of $C-J$ such that $\overline{N}_3 \cdot J$ contains p, q, r, and since $\overline{N}_1 \cdot J$ contains p, q, r, the circuit J would not be a split circuit (definition). Now any two points of p, q, r lie together on the same c.d.b. of C within $\overline{R}_3$. For if every arc from p to q (or to r) lying in R_3 contained a point of C, there would exist, as above, a connected subset of C lying in $\overline{R}_3$ and having points on $l+m+n$ that separate p and q on $l+m+n$. There would then exist a component N_4 of $C-J$ such that $\overline{N}_4 \cdot J$ contains p, q, and r; but this is impossible. Now let the circuit J_1 above be the boundary of a domain r_1 complementary to C and lying in R_2. Let l be the arc (of the three l, m, n) which lies on $J_1 \supset p+q$. Then since R_3 contains no points of $J_1+J_2+J_3$, the domain r_1 is a subset of R_2, but not of R_3, and has p and q on its boundary. Let r_2 be a domain complementary to C which is a subset of R_3 and has p and q on its boundary. Now R_3 is not a complementary domain of C since we are assuming that p, q, r

* V. W. Adkisson, *Cyclicly connected continuous curves whose complementary domain boundaries are homeomorphic, preserving branch points*, Comptes Rendus des Séances de la Société des Sciences et des Lettres de Varsovie, Class 3, vol. 23 (1930), p. 167, Theorem 2. If M is a cyclicly connected continuous curve lying on a sphere S, and T is a continuous (1-1) correspondence such that $T(M)=M$, a necessary and sufficient condition that T be extendable to S is that for every boundary J of a complementary domain of M, $T(J)$ be also the boundary of a complementary domain of M.

† C. Kuratowski, *Sur le problème des courbes gauches en topologie*, Fundamenta Mathematicae, vol. 15 (1930), p. 274, Lemma III′.

do not all lie on the same c.d.b. in $\overline{R}_2$. The above process can then be repeated and a third complementary domain r_3 obtained which is different from r_1 and r_2 but also has p and q on its boundary. This process may be continued indefinitely. Hence there exists an infinite number of domains $r_1, r_2, r_3, \cdots$ each complementary to C and having p and q on its boundary. But this is impossible since there is at most a finite number of complementary domains of C of diameter greater than any $\epsilon>0$,* and the diameter of each r_i is equal to or greater than the distance between p and q. We conclude that there is a c.d.b. L within $\overline{R}_2$ containing p, q, r.

Now any fourth point s of $\overline{N}_1 \cdot J$ must also lie on L. For s and one of the three points p, q, r, say r, must separate the other two, p and q, on J. Since r and s must lie on the same c.d.b. of C in $\overline{R}_2$ there exists an arc $\langle rs \rangle$ in R_2 such that $\langle rs \rangle \cdot C = 0$, and in like manner an arc $\langle pq \rangle$ in R_2 such that $\langle pq \rangle \cdot C = 0$. But $\langle rs \rangle$ and $\langle pq \rangle$ must then have a common point and hence lie in a common complementary domain of C. Therefore p, q, r, and s all lie on L, and all points of $\overline{N}_1 \cdot J$ must lie on L.

Let D_1 be the complementary domain of C bounded by L, and let (N) represent the set of all components N_i of $C-J$ in R_1 for which $\overline{N}_i \cdot J = \overline{N}_i \cdot L$. Let (N') be a topological image of (N) in D_1 such that $C-(N)+(N')$ is a map of C. This is always possible since obviously it would be possible if C were mapped so that L is a circle. This process can be repeated on the map $C-(N)+(N')$ so that finally a map C' is obtained in which R_1 is a complementary domain of C'. It follows by well known methods in analysis situs that C' is a topological map of C. For if the above process is necessary an infinite number of times,† it involves complementary domains $D_1, D_2, D_3, \cdots$ of C of which only a finite number are of diameter greater than any $\epsilon>0$.

By the same method as above it is possible to map C so that the image of the split circuit J is not a c.d.b. of the image of C since there are at least two components of $C-J$.

Now let C' be a map of C on S' in which the image of J is a c.d.b. of C' and C'' a map on S'' in which the image of J is not a c.d.b. of C''. Then every homeomorphism of C' into C'' is not extendable to S' and S'' since complementary domain boundaries are not preserved. Hence C does not have a unique map, and the theorem is proved.

Corollaries 1 and 2 follow directly from the preceding proof.

* Schoenflies proved (1908) that the complementary domains of a continuous curve are countable, and that at most a finite number have diameters greater than any $\epsilon>0$. See R. L. Moore, *Report on continuous curves from the viewpoint of analysis situs*, Bulletin of the American Mathematical Society, vol. 29 (1923), pp. 290, 295.

† The components N_i are countable. See R. L. Wilder, *Concerning continuous curves*, Fundamenta Mathematicae, vol. 7 (1925), p. 360, Theorem 9.

Proof of Theorem III. We shall assume that M has a split point and at the same time a unique map and show that this leads to a contradiction.

Let M be a map on the sphere S, p a split point of M, and A and B two closed connected subsets of M such that $M=A+B$, $A\cdot B=p$, and such that if A or B is an arc, p is not an end point of this arc. Let R be a complementary domain of M whose boundary contains points of both A and B. Let $x\neq p$ be a point of A in the boundary of R and $z\neq p$ a point of B in the boundary of R. Let $\langle xyz\rangle$ be an arc in R, xpz an arc of M, $A\supset$ arc xp, and $B\supset$ arc pz. Let A_1 be a connected component of $A-xp$ such that $\overline{A}_1$ has a point r on $xp-x$, and B_1 a connected component of $B-pz$ such that $\overline{B}_1$ has a point s on $pz-z$. Since neither A nor B is a simple arc with end point at p, it is possible to choose x and z so that this latter condition may be satisfied. There are two cases to consider.

Case 1. $r\neq s$. Let M' represent a map of M on the sphere S'. Let T be a homeomorphism such that $T(M)=M'$ and $T(S)=S'$. This is possible since we assume that M has a unique map. Throughout this proof a primed set will indicate the topological image under T of the unprimed set. Now either the set A_1 lies in the region D_1 of S bounded by the simple closed curve $C=xpzyx$ while B_1 lies in the other region D_2 bounded by C, or A_1 and B_1 lie in the same region, say D_1. We shall assume the latter, but the proof is practically the same in either case.

Since T is extendable to S and S', the sets A_1' and B_1' lie in D_1'. We now construct a new map M'' of M on S' as follows: Let H_i', $(i=1,2)$, be the subset of $A'-$arc $x'p'$ that lies on D_i'. The set H_i' includes A_i'. Let H_1'' be a topological image of H_1' lying in $D_2'-B'\cdot D_2'$, and H_2'' a topological image of H_2' lying in $D_1'-B'\cdot D_1'$ such that $x'p'+H_1''+H_2''$ is a topological image of A'. Then $M''=x'p'+H_1''+H_2''+B'$. Let U be a homeomorphism such that $U(M)=M''$ and $U(S)=S'$. Let $U(xyz)=x'y''z'$, $U(H_i)=H_i''$, and for points in $B+xpz$ let $U=T$. Since U is extendable, $U(D_1)=D_1''$ is a region of S' containing $U(M\cdot D_1)$. Let f be any arc of S' from a point of A_1'' (the topological image of A_1' in H_1'') to a point of B_1' which lies entirely in D_1'' including end points. Such an arc must intersect C', the boundary of D_1', since B_1' lies in D_1' and A_1'' lies outside D_1'; and since $x'p'z'$ is in the boundary of D_1'' the arc f must intersect $\langle x'y'z'\rangle$. Then any arc g from r' to s' lying, except for end points, in D_1'' must intersect $\langle x'y'z'\rangle$. To show this let d and e be regions about r' and s' respectively, that do not contain points of $x'y'z'$. It is possible to obtain an arc t that lies entirely in d and joins a point of $g-r'$ to a point of A_1''. In like manner we obtain an arc u in e joining a point of $g-s'$ to a point of B_1'. The arcs t and u plus the proper subset of g then yield an arc h from A_1'' to B_1' lying entirely in D_1''. But h must then intersect $\langle x'y'z'\rangle$, and

since neither t nor u can intersect $\langle x'y'z'\rangle$ the arc g must have a point in common with $\langle x'y'z'\rangle$. Hence there must be a connected subset of $x'y'z'$ lying in D_1'' with two points on $C''=x'y''z'p'x'$ (the boundary of D_1'') that separate r' and s' on C''.* One of these points must obviously lie on the arc $r's'$ of C'' where $r's' \subset \langle x'p'z'\rangle$. But this is impossible since $\langle x'p'z'\rangle$ and $\langle x'y'z'\rangle$ have no common points. Therefore the assumption that every homeomorphism of M is extendable has led to a contradiction, and we conclude that M has no unique map.

Case 2. $r=s=p$. We use the same notation as in Case 1 and obtain in the same manner the map M''. Any arc joining a point of A_1'' to a point of B_1' and lying entirely in D_1'' must intersect $\langle x'y'z'\rangle$. Let d be a region about p' that contains no points of $x'y'z'$. Let f be an arc from A_1'' to B_1' lying in D_1'' and also in d. Then f cannot intersect $\langle x'y'z'\rangle$ but must intersect $\langle x'y'z'\rangle$. This contradiction shows that M has no unique map and completes the proof of Theorem III.

Proof of Theorem IV. First, assume that M has a unique map. If M is acyclic it must consist of either a simple continuous arc or a single triod since M contains no split point (Theorem III). Obviously these are the only two acyclic peanian continua without split-points.

If M is not acyclic it contains at least one cyclic element C which is a maximal cyclic curve of M. Since M cannot contain a split-point there is only one such cyclic element C. For if there were a second cyclic element C_1 which is a maximal cyclic curve of M, there would exist a cut-point p of M separating C and C_1,† and obviously p would also be a split-point.

If $M-C$ contains a maximal connected acyclic subset with a branch-point p, or if $M-C$ contains two arcs with a common end point p on C, then p is in either case a split-point. Therefore, $M-C$ contains at most a countable number of simple arcs‡ with distinct end points on C.

We shall now assume that C is not triply connected and obtain a contradiction of the assumption that M has a unique map. If C is not triply connected, C contains a split circuit J (Theorem I). Let M be a map of M on the sphere S and assume that the components N_1 and N_2 of $C-J$ lie in the regions R_1 and R_2 of S bounded by J. There exists a c.d.b. L of C lying in $\overline{R}_2$ such that $\overline{N}_1 \cdot J = \overline{N}_1 \cdot L$ (this was shown in the proof of Theorem II). Let D be the complementary domain of C bounded by L. Let $\sum a_i$ be the set of arcs of M lying in D with end points in $\overline{N}_1 \cdot L$ and $\sum b_i$ the set of arcs of $M \cdot D$ not included in $\sum a_i$. We now obtain a new map M' of M on S as follows: Let

* Kuratowski, loc. cit., Lemma III'.

† See G. T. Whyburn, loc. cit., p. 168.

‡ R. L. Wilder, loc. cit., p. 360, Theorem 9.

(K) represent all arcs of $M-C$ with end points on N_1, and let N_1' be a map in D of $N_1+(K)$ such that $\overline{N}_1 \cdot L = \overline{N}_1' \cdot L$ and $C-N_1+N_1'$ is a topological image of $C+(K)$. Let D_1 be the complementary domain of $M-[N_1+(K)]$ that contains $N_1+(K)$, and let $\sum a_i'$ be a topological image of $\sum a_i$ in D_1 such that each arc $\bar{a}_i'$ has one end point on L, and the set of all these end points of $\sum \bar{a}_i'$ is identical with the set of end points of $\sum \bar{a}_i$ on L. For each b_i we obtain a new arc b_i' as follows: If $\bar{b}_i \cdot N_1' = 0$, then $b_i' = b_i$. If $\bar{b}_i \cdot N_1' \neq 0$, let d_i be a region about Q_i (the end point of b_i on L) containing no point of N_1'. Then b_i' is taken as a simple arc which is a subset of b_i and lies in d_i with end point Q_i. Then $M' = M - N_1 - (K) + N_1' - \sum a_i + \sum a_i' - \sum b_i + \sum b_i'$ is a topological image of M. But every homeomorphism of M into M' cannot be extended to S since L is a c.d.b. of M but not a c.d.b. of M'. Hence M does not have a unique map, and this contradiction leads to the conclusion that C must be triply connected.

Now suppose C does not consist of a simple closed curve, and there is an arc b in $M-C$ with end point p on two bounding circuits, J and L, of C. Let $(M-b)$ be any map of $(M-b)$ on S. Then J and L bound two complementary domains of C and are outer boundaries of two complementary domains D_1 and D_2 of $M-b$. Let M' be a map of M obtained by mapping b in D_1 with end point at p, and M'' a map of M obtained by mapping b in D_2 with end point at p. We have now two essentially different maps of M. Hence the end points of the arcs in $M-C$ that lie in C must lie in one, and only one, bounding circuit of C.

If C consists of a simple closed curve, $M-C$ cannot contain more than one arc. The proof of this statement is not difficult and will be omitted.

If M consists of a simple continuous arc, the sufficiency of condition (1) follows from the fact that any homeomorphism between two arcs is extendable to a homeomorphism of their planes.* If M consists of a simple closed curve plus an arc, the proof that M has a unique map is easily obtained from the Schoenflies theorem that any homeomorphism between two simple closed curves can be extended to a homeomorphism of their planes.

The proof that a triod has a unique map may also be obtained by a simple application of the Schoenflies theorem.

Suppose C is not a simple closed curve. Let M be a map of M on S, and M' a map on S'. The map of C is unique (Theorem II), and from Corollary 1 we see that a c.d.b. of C in one map has an image which is a c.d.b. in every map of C. Let $\sum a_i$ be the arcs of $M-C$ with end points on the same bounding cir-

* R. L. Moore, *Conditions under which one of two given closed linear point sets may be thrown into the other one by a continuous transformation of a plane into itself*, American Journal of Mathematics, vol. 48 (1926), p. 67.

cuit J of C. Then in the map of C on S these arcs lie in the same complementary domain D of C bounded by J, and the map $\sum a_i'$ of $\sum a_i$ on S' must lie in the complementary domain D' of C' bounded by J' (the image of J) since the end points of $\sum \bar{a}_i'$ lie on J' and on no other bounding circuit of C'.

Now any homeomorphism $T(M)=M'$ can be extended to D and D'. Let pq be one of the arcs $\bar{a}_i$ in $\bar{D}$ with end point p on J such that either pq or its image $p'q'$ on S' is of diameter greater than some $\epsilon>0$. Let $\langle qxy\rangle$ be an arc in $D-\sum a_i$ where y is a point of J but not an end point of some arc a_i in D. Let $\langle q'x'y'\rangle$ be an arc in $D'-\sum a_i'$ where $y'=T(y)$. Let r and s be points of J that separate p and y, and let $r'=T(r)$ and $s'=T(s)$. Let a_k be any arc of $\sum a_i$ that lies in the subset d of D bounded by $pqxyrp$. Then a_k' must lie in the subset d' of D' bounded by $p'q'x'y'r'p'$. For if a_k' lies in $D'-\bar{d}'$, it must have an end point at either p' or y'. But this is impossible since y was selected not to be an end point, and p cannot be an end point common to two arcs in D. Hence if $\sum b_i$ is the subset of $\sum a_i$ lying in d, then $T(\sum b_i)$ is the subset of $\sum a_i'$ lying in d'. The same would be true of any similar subdivision of D and D'. In fact *sides are preserved under* T as used by Gehman.* Therefore T can be extended to D and D', and in like manner to each complementary domain of C and C', and finally to S and S'. The proof would necessarily be a partial duplication of Gehman's proof and is omitted. This completes the proof of Theorem IV.

Proof of Theorem V. Cases (1), (2), (3) of Theorem V follow easily from Theorem IV.

If M has a unique map and is not acyclic or not a simple closed curve, Theorem III shows that M contains one, and only one, cyclic element C which is a maximal connected cyclic curve of M. Furthermore C cannot contain two distinct bounding circuits or consist of a single circuit which would be the outer boundary of two complementary domains of M in any map. For if this were true there would be a circuit J which would be the outer boundary of a bounded complementary domain R of M in some map on a plane S (Corollary 1). Now let M' be a map of M on a plane S' in which the image of the boundary of R is the boundary of the unbounded complementary domain R' of M'. Then obviously any homeomorphism of M into M' which carries the boundary of R into the boundary of R' cannot be extended to R and R'. Therefore C contains only one bounding circuit, and Claytor's result (Corollary (C), p. 810) shows that C consists of a closed 2-cell.

Since M cannot contain a split point, $M-C$ can consist of at most a count-

* H. M. Gehman, *On extending a continuous* (1-1) *correspondence of two plane continuous curves to a correspondence of their planes*, these Transactions, vol. 28 (1926), proof of Theorem I, pp. 256–260.

able number of distinct arcs with distinct end points on the bounding circuit of C.

Now suppose M consists of the curve (4) in the theorem. The arcs $M-C$ must all lie in the unbounded complementary domain of C in any map of M in the plane. Let D and D' be the two unbounded complementary domains of two maps of M and M' respectively. Then any homeomorphism of M into M' can be extended to D and D' as indicated in the last paragraph of the proof of Theorem IV.

A point p of a peanian continuum M is said to be a *local cut-point* of M if, and only if, p is a cut-point of some connected open subset of M.*

The following theorem will be stated without proof:

THEOREM VI. *If M is a peanian continuum in a plane S, any non-cut-point p of M lies on two (or more) complementary domain boundaries of M if, and only if, p is a local cut-point of M.*†

Now if M is a plane peanian continuum and C a maximal cyclic curve of M, every point in $C\cdot(M-C)$ must lie on a bounding circuit of C. Since C contains no cut-point (of C) we can make the following alternative statement of Theorem IV:

The plane peanian continuum M has a unique map on the sphere if, and only if, one of the following conditions holds:

(1) *M is acyclic and consists of either a simple arc or a triod,*

(2) *M contains one cyclic element C which is a maximal triply connected cyclic curve of M, and $M-C$ consists of at most a countable number of arcs, $a_1, a_2, a_3, \cdots$, such that $\bar{a}_i\cdot\bar{a}_j=0$, $(i\neq j)$, and each $\bar{a}_i\cdot C$ is a non-local separating point of C, provided that if C is a simple closed curve, then $M-C$ is at most a simple arc.*

---

* See G. T. Whyburn, *Local separating points of continua*, Monatshefte für Mathematik und Physik, vol. 36 (1929), pp. 305–314.

† This theorem is well known to topologists although the author has been unable to find it stated explicitly in any published paper. See, however, G. T. Whyburn, *Local separating points of continua*, loc. cit., Theorem 6, and G. T. Whyburn, *Concerning points of continuous curves defined by certain im kleinen properties*, Mathematische Annalen, vol. 102 (1929), pp. 313–336, Theorem 31.

UNIVERSITY OF ARKANSAS,
FAYETTEVILLE, ARK.

THE LAW OF APPARITION OF PRIMES IN A LUCASIAN SEQUENCE*

BY
MORGAN WARD

I. Introduction

1. We call a sequence of rational integers

$$(u): \quad u_0, u_1, u_2, \cdots, u_n, \cdots$$

Lucasian (Ward [1]†) if it satisfies a linear recursion relation with constant integral coefficients, and if u_n divides u_m whenever n divides m. The adjective "Lucasian" is chosen in honor of the French mathematician Eduard Lucas who first developed a theory of these sequences‡ (Lucas [1], [2]). We are concerned here with the fundamental problem of determining a priori all the terms of such a sequence divisible by any preassigned modulus m.

Call the suffix k of a term u_k of (u) divisible by m a place of apparition of m in (u), and let $\mathfrak{S}_m$ denote the set of all places of apparition of m. It follows from the results established in Ward [1] that the set $\mathfrak{S}_m$ consists in general§ of all multiples of a finite number of places of apparition $\rho_1, \rho_2, \cdots, \rho_s$ called the ranks of apparition of m in (u) with the defining properties

$$u_\rho \equiv 0 \pmod{m}, \qquad u_s \not\equiv 0 \pmod{m} \text{ if } s \text{ divides } \rho.$$

The least common multiple‖ $\rho = [\rho_1, \rho_2, \cdots, \rho_s]$ of the ranks of apparition of m in (u) is called simply the *rank* of m in (u). The places of apparition of m in (u) are periodic modulo ρ, and ρ divides the restricted period¶ of (u) modulo m. Furthermore if $m = a \cdot b$ where a and b are co-prime, then the set $\mathfrak{S}_m$ of places of apparition of m is the cross cut of the sets $\mathfrak{S}_a$ and $\mathfrak{S}_b$, and each rank of apparition of m is the least common multiple of ranks of apparition of a and b.

Our fundamental problem reduces then to determining the ranks of ap-

* Presented to the Society, February 26, 1938; received by the editors July 13, 1937.

† The numbers [1], [2], ··· refer to the bibliography at the close of the paper.

‡ Lucas confined himself in the main to the case when the recursion relation is of order two.

§ An exception occurs only if m divides every term of (u) beyond a certain point.

‖ We use $[a, b, \cdots]$ and $(a, b, \cdots)$ to denote the least common multiple and greatest common divisor of the integers $a, b, \cdots$.

¶ The restricted period of (u) modulo m is the least positive integer μ such that $u_{n+\mu} \equiv a u_n \pmod{m}$ for all large n, where a is a constant integer. For the terminology of the theory of recurring series which we employ, see Ward [2].

parition of primes and powers of primes in (u). In the terminology of Lucas,* we must discover the "law of apparition" of primes in (u), and the "law of repetition" of primes in (u). I shall confine myself here to the first problem; the modulus m will invariably be a prime number p.

2. It will be well at this point to exhibit some Lucasian sequences. Let

$$f(x) = x^k - c_1x^{k-1} - \cdots - c_k, \qquad g(x) = x^l - d_1x^{l-1} - \cdots - d_l$$

be two polynomials with rational integral coefficients $c_1, \cdots, d_l$. For simplicity of exposition we assume that $f(x)$ and $g(x)$ have non-vanishing discriminants and resultant.† Let $\alpha_1, \cdots, \alpha_k; \beta_1, \cdots, \beta_l$ denote the roots of $f(x)=0$ and $g(x)=0$ respectively. Then none of the $k(2l+k-1)/2$ differences $\alpha_i-\beta_j$, $\alpha_i-\alpha_r$, $r\neq i$, vanish.

Consider now the sequences (U): $U_0, U_1, \cdots$; (R): $R_0, R_1, \cdots$, where

$$U_n = U_n(f) = \prod_{i<r}\left(\frac{\alpha_i^n - \alpha_r^n}{\alpha_i - \alpha_r}\right), \qquad R_n = R_n(f, g) = \prod\left(\frac{\alpha_i^n - \beta_j^n}{\alpha_i - \beta_j}\right).$$

Then U_n and R_n are rational integers, and both sequences are clearly divisibility sequences. Both sequences are also linear (Ward [1]). Hence, *both sequences are Lucasian.* The sequence U_n for $k=2$ is the classical Lucas function (Lucas [1]), while R_n for $g(x)=x-1$ is equivalent to the function studied by T. A. Pierce [1], P. Poulet [1], and D. H. Lehmer [2].‡

We shall call the polynomials $f(x)$ and $g(x)$ the *generators* of (R) and (U). We refer to both types of sequences as R-sequences.

The determination of the law of apparition for R-sequences is of particular importance because it appears probable that *all* Lucasian sequences may be exhibited as R-sequences or divisors of R-sequences.§ (See next section.) I shall show here in detail that the determination of the law of apparition

---

* See Lucas [1], pp. 209, 289, 294, or Lehmer [1], pp. 421, 422.

† This restriction is removed in the body of the paper.

‡ It is possible to exhibit both (R) sequences and (U) sequences as Pierce sequences. For if we let $\bar\beta=\beta^{-1}$, then $(\alpha^n-\beta^n)/(\alpha-\beta)=\beta^{n-1}[(\alpha\bar\beta)^n-1]/(\alpha\bar\beta-1)$. Accordingly if we denote the kl products $\alpha_i\bar\beta_j$ in any order by $\epsilon_1, \epsilon_2, \cdots, \epsilon_{kl}$, then

$$R_n = (-1)^{kl(n-1)}d_l^{k(n-1)}\prod_{h=1}^{kl}\left(\frac{\epsilon_h^n - 1}{\epsilon_h - 1}\right),$$

and $(\epsilon_1^n-1)(\epsilon_2^n-1)\cdots(\epsilon_{kl}^n-1)$ is the function studied by Pierce in the paper cited.

A similar result holds for (U). Since we must then deduce the properties of (R) from a polynomial $(x-\epsilon_1)\cdots(x-\epsilon_{kl})$ of higher degree than $f(x)$ or $g(x)$ with non-integral coefficients whose factorization depends in a highly complicated manner upon $f(x)$ and $g(x)$, the reduction appears to be of only formal interest.

§ With the qualifications described in §3, I have found empirically no Lucasian sequences which are not R-sequences.

depends upon the fundamental problem of determining the period of a mark in a finite field. My results are sufficiently precise to give a good deal of specific information about the terms divisible by a given prime in any numerical example of an R-sequence.

The sequence (U) is also of importance because of the following theorem:

THEOREM 2.1. *Let the Lucasian sequence (u) belong to the polynomial $f(x)$, and let p be any prime which does not divide the discriminant of $f(x)$. Then every place of apparition of p in (u) is a place of apparition of p in the Lucasian sequence (U) generated by $f(x)$.*

3. Another extensive class of Lucasian sequences arises as follows. Consider for simplicity a sequence (U) with an irreducible generator $f(x)$. The galois group of $f(x)$ may be represented as a transitive permutation group upon the roots $\{\alpha_1\}, \{\alpha_2\}, \cdots, \{\alpha_k\}$.

Now let us represent the group as a permutation group upon the $k(k-1)/2$ *pairs* of roots $\{\alpha_1, \alpha_2\}, \{\alpha_1, \alpha_3\}, \cdots, \{\alpha_{k-1}, \alpha_k\}$.

If the group is singly transitive over the $\{\alpha_i\}$, the pairs $\{\alpha_i, \alpha_j\}$ may be separated into $\kappa \geqq 2$ transitive sets

$$\{\alpha_{i_1}, \alpha_{i_1'}\}, \{\alpha_{i_2}, \alpha_{i_2'}\}, \cdots, \{\alpha_{i_{s_i}}, \alpha_{i_{s_i}'}\}$$
$$i = 1, 2, \cdots, \kappa; s_1 + s_2 + \cdots + s_K = k(k-1)/2.$$

We have a corresponding arithmetical factorization of the general term U_n of (U) into a product of κ rational integers:

$$U_n = \prod_{i=1}^{K} U_n^{(i)}, \qquad U_n^{(i)} = \prod \left(\frac{\alpha_{i_1}^n - \alpha_{i_1'}^n}{\alpha_{i_i} - \alpha_{i_1'}} \right).$$

Each of the κ sequences $(U^{(i)})$ is obviously Lucasian.*

We shall refer to sequences obtained in this manner as *divisors* of R-sequences. The determination of the laws of apparition of primes in divisors of

* For example, suppose that $k=4$ and that $f(x) = x^4 - c_1x^3 - c_2x^2 - c_3x - c_4 \equiv x^4 + (2Q-R)x^2 + Q^2$ where Q and R are co-prime integers and R is not a square. Then with a proper notation, $(x-\alpha_1)(x-\alpha_2) = x^2 - R^{1/2}x + Q$, $\alpha_3 = -\alpha_1$, $\alpha_4 = -\alpha_2$. There are two transitive sets of the $\{\alpha_i, \alpha_j\}$; namely, $\{\alpha_1, \alpha_2\}$, $\{\alpha_1, \alpha_4\}$, $\{\alpha_2, \alpha_3\}$, $\{\alpha_3, \alpha_4\}$ and $\{\alpha_1, \alpha_3\}$, $\{\alpha_2, \alpha_4\}$.

We find that $U_n = U_n^{(1)} U_n^{(2)}$ where

$$U_n^{(1)} = \left(\frac{\alpha_1^n - \alpha_2^n}{\alpha_1 - \alpha_2} \right)^2 \left(\frac{\alpha_1^n - (-\alpha_2)^n}{\alpha_1 + \alpha_2} \right)^2, \qquad U_n^{(2)} = (4\alpha_1\alpha_2)^{n-1} = (4Q)^{n-1}.$$

Now $(\alpha_1^n - \alpha_2^n)/(\alpha_1 - \alpha_2)$ is one of the important functions introduced by D. H. Lehmer in his doctor's thesis (Lehmer [1]), and $[\alpha_1^n - (-\alpha_2)^n]/(\alpha_1 + \alpha_2)$ is immediately expressible in terms of Lehmer's U_n and V_n.

The function $N(\alpha^n - \beta^n)$ studied by Marshall Hall (Hall [2]) may be similarly exhibited as a divisor of a certain R-sequence.

R-sequences is an important part of our general problem. But to avoid stretching the present paper to an inordinate length, we shall give our investigations elsewhere. The problem amounts to correlating the results obtained in this paper by the use of Schatanovski's principle (§7) with results obtained from the Dedekind-Hilbert theory of the ideals of a galois field.

4. The law of apparition of primes in R-sequences is determined as follows. Consider first the sequence (R). We show (§§6, 7) that it suffices to consider primes which do not divide the resultant of the generators of (R). We have decompositions of $f(x)$ and $g(x)$ modulo p of the form

$$f(x) \equiv f_1(x)^{a_1} \cdots f_r(x)^{a_r};\ g(x) \equiv g_1(x)^{b_1} \cdots g_s(x)^{b_s} \pmod{p},$$

where the polynomials f_i and g_j are primary, irreducible and co-prime in pairs modulo p. We show in §8 that we have a corresponding decomposition of the general term of (R) modulo p

$$R_n(f, g) \equiv \prod_{i,j} \{R_n(f_i, g_j)\}^{a_i b_j} \pmod{p}.$$

In the terminology of Ward [1], the sequence (R) factors modulo p into a product of simpler sequences; for the f_i and g_j are irreducible modulo p. But then (Ward [1]) the set $\mathfrak{S}_p$ of places of apparition of p in (R) is the union of the sets of places of apparition of p in the sequences $(R(f_i, g_j))$. Therefore *in discussing the law of apparition of primes in* (R) *we may assume that the generators of* (R) *are irreducible modulo* p. A like simplification holds for the sequence (U) (§9).

5. If the generator of (U) is irreducible modulo p, the law of apparition of p in (U) takes the following beautifully simple form, affording a far-reaching generalization of the classical results of Lucas [1]:

THEOREM 5.1. *Let* $f(x)$ *be irreducible modulo* p, *and let* λ *be its period* modulo* p. *Let* $k = q_1^{c_1} q_2^{c_2} \cdots q_K^{c_K}$ *be the decomposition of its degree* k *into prime factors. Let* $\rho(s)$ *be defined for any positive integer* s *as the residual† of* $p^s - 1$ *with respect to* λ; *that is, the quotient of* λ *by the greatest common divisor of* λ *and* $p^s - 1$. *Then the ranks of apparition of* p *in* (U) *occur among‡ the* κ *numbers* $\rho(k/q_1), \cdots, \rho(k/q_K)$, *the rank of* p *in* (U) *divides* $\rho(k/q_1 q_2 \cdots q_K)$, *and* p *has at most* κ *ranks of apparition.*

We observe that the numbers $\rho(k/q)$ are known as soon as the period is known.

* The period of $f(x)$ modulo p is by definition the smallest positive value of λ such that $x^\lambda \equiv 1 \pmod{p, f(x)}$.

† The operation of residuation has important arithmetical applications. I have developed some of these in the paper, Ward [3], which arose out of the present investigations.

‡ We must exclude from the set of $\rho(k/q)$ any element which is a multiple of any other.

Unlike the ranks of apparition of p in (U), the ranks of apparition of p in (R) are not obtainable from the periods of the generators $f(x)$ and $g(x)$ of (R) alone when the generators are irreducible modulo p. If $f(\alpha)=0$, $g(\beta)=0$, the ranks of apparition occur among the l periods $\sigma_1, \sigma_2, \cdots, \sigma_l$ modulo p of the algebraic numbers $\alpha\beta^{-1}, \alpha\beta^{-p}, \cdots, \alpha\beta^{-p^{l-1}}$ in the galois field of the roots of the generators (§11). In §14, we assign upper and lower limits to the periods σ in terms of the periods and restricted periods of $f(x)$ and $g(x)$.

The least common multiples of pairs of the periods σ have the following remarkable property (§13):

$$[\sigma_s, \sigma_t] = [\sigma_s, \sigma_{s\pm(m,t-s)}].$$

Here m is the least common multiple of the degrees of $f(x)$ and $g(x)$ and we adopt the convention that $\sigma_x=\sigma_y$ if $x\equiv y \pmod{l}$.

It appears unlikely that results of simplicity comparable to Theorem 5.1 exist for the law of apparition of primes in (R).

II. Reduction to R-sequences with irreducible generators

6. This section is devoted to some algebraic preliminaries. Let x; $y_1, y_2, \cdots, y_k$; $z_1, z_2, \cdots, z_l$ be $k+l+1$ indeterminates, and let $Y_1, -Y_2, \cdots, (-1)^{k-1}Y_k$; $Z_1, -Z_2, \cdots, (-1)^{l-1}Z_l$ be the $k+l$ elementary symmetric functions of the indeterminantes y, z defined by*

$$(x-y_1)(x-y_2)\cdots(x-y_k) = x^k - Y_1x^{k-1} - \cdots - Y_k,$$
$$(x-z_1)(x-z_2)\cdots(x-z_l) = x^l - Z_1x^{l-1} - \cdots - Z_l.$$

By the fundamental theorem on symmetric functions, the polynomials

$$(6.1)\quad \Theta_{k,l}(y,z) = \prod_{i=1}^{k}\prod_{j=1}^{l}\left(\frac{y_i^n - z_j^n}{y_i - z_j}\right), \qquad \Psi_k(y) = \prod_{\substack{i,j=1\\ i<j}}^{k}\left(\frac{y_i^n - y_j^n}{y_i - y_j}\right)$$

may be expressed as polynomials in the Y and Z with integral coefficients; we write

$$(6.2)\qquad \Theta_{k,l}(y,z) = P_{k,l}(Y,Z), \qquad \Psi_k(y) = Q_k(Y).$$

Suppose now that $t_1, t_2, \cdots, t_m$ are m new indeterminates where $k\geqq m\geqq 1$, $l\geqq m\geqq 1$, and consider the effect of substituting t_1 for y_k and z_l, t_2 for y_{k-1} and z_{l-1}, t_3 for y_{k-2} and z_{l-2}, and so on, in the identity (6.2). If we let

$$(x-y_1)\cdots(x-y_{k-m}) = x^{k-m} - Y_1'x^{k-m-1} - \cdots - Y_{k-m}',$$
$$(x-z_1)\cdots(x-z_{l-m}) = x^{l-m} - Z_1'x^{l-m-1} - \cdots - Z_{l-m}',$$
$$(x-t_1)\cdots(x-t_m) = x^m - T_1'x^{m-1} - \cdots - T_m',$$

* Minus signs are introduced so that the associated difference equation used later $\Omega_{n+k}=Y_1\Omega_{n+k-1}+\cdots+Y_k\Omega_n$ may have all its signs positive.

then the polynomial $P_{k,l}$ on the right of (6.2) is transformed into a polynomial $P^*_{k,l,m}$ in the arguments Y', Z', T' with integral coefficients. Its expression in terms of y, z, t is easily found to be

$$n^m T_m'^{-1}\Theta_{k-m,l-m}(y, z)\Theta_{k-m,m}(y, t)\Theta_{m,l-m}(t, z)\Psi_m^2(t).$$

Hence by (6.2)

$$P^*_{k,l,m}(Y', Z', T') = n^m T_m'^{n-1} P_{k-m,l-m}(Y', Z') P_{k-m,m}(Y', T') P_{m,l-m}(T', Z') Q_m^2(T'). \tag{6.3}$$

Now let $R_n = P_{kl}(Y, Z)$, $U_n = Q_k(Y)$, $R_n^* = P^*_{k,l,m}(Y', Z', T')$, and consider the sequences

$$(R): \quad R_0, R_1, R_2, \cdots,$$
$$(U): \quad U_0, U_1, U_2, \cdots,$$
$$(R^*): \quad R_0^*, R_1^*, R_2^*, \cdots.$$

THEOREM 6.1. *(R), (U), and (R^*) are Lucasian in the rings formed by adjoining respectively Y, Z; Y; Y', Z', T' to the ring of rational integers.*

Proof. The sequences evidently lie in the specified rings. Consider (R). Since its general term is a product of cyclotomic functions $(y^n - z^n)/(y - z)$ having the divisibility property, (R) has the same property; that is, R_n divides R_m if n divides m, and the division may be performed in the ring of Y and Z. The linearity of (R) over the ring follows from a general theorem in Ward [1]. Hence (R) is Lucasian. Similarly (U) is Lucasian. Then (R^*) as a product of the seven Lucasian sequences with general terms n^m, $T_m'^{n-1}$, $P_{k-m,l-m}(Y', Z')$, $P_{k-m,m}(Y', T')$, $P_{m,l-m}(T', Z')$, $Q_m(T')$, $Q_m(T')$, and is also Lucasian (Ward [1]).

7. We now consider the sequence (R) of §2 of the introduction. Let $\mathfrak{R}$ denote the ring of rational integers, and let

$$f(x) = x^k - c_1 x^{k-1} - \cdots - c_k; \qquad g(x) = x^l - d_1 x^{l-1} - \cdots - d_l$$

be two polynomials with fixed rational integral coefficients. Let $\alpha_1, \cdots, \alpha_k$; $\beta_1, \cdots, \beta_l$ be their roots, D_f and D_g their discriminants, and

$$R_{f,g} = \pm \prod (\alpha_i - \beta_j)$$

their resultant. If $R_{f,g}$ does not vanish, we define a sequence

$$(R): \quad R_0, R_1, R_2, \cdots,$$

in the notation of §6 by $R_n = \Theta_{k,l}(\alpha, \beta) = P_{k,l}(c, d)$.

If $R_{f,g}$ vanishes, then

$$f(x) = f'(x)h'(x), \qquad g(x) = g'(x)h'(x), \tag{7.1}$$

where

$$\begin{aligned} f'(x) &= x^{k-m} - c_1' x^{k-m-1} - \cdots - c_{k-m}', \\ g'(x) &= x^{l-m} - d_1' x^{l-m-1} - \cdots - d_{l-m}', \\ h'(x) &= x^{m} - e_1' x^{m-1} - \cdots - e_m', \end{aligned} \tag{7.2}$$

$$k \geqq m \geqq 1; l \geqq m \geqq 1,$$

are polynomials in $\mathfrak{R}$ and $R_{f',g'} \neq 0$. Deviating for simplicity from the notation of the previous section, we now define the sequence (R) (instead of a new sequence (R^*)) by letting $R_n = P^*_{k,l,m}(c', d', e')$. In each case we obtain a Lucasian sequence over $\mathfrak{R}$.

Consider now the places of apparition of any prime number p in (R). There are two cases to consider according as p does or does not divide the resultant $R_{f,g}$.

Case 1. $R_{f,g} \not\equiv 0 \pmod{p}$. Then $R_n \equiv 0 \pmod{p}$ if and only if

$$\Theta_{k,l}(\alpha, \beta) = \prod \left(\frac{\alpha_i^n - \beta_j^n}{\alpha_i - \beta_j} \right) \equiv 0 \pmod{p}.$$

Case 2. $R_{f,g} \equiv 0 \pmod{p}$. In this case (7.1) and (7.2) hold modulo p with $R_{f',g'} \not\equiv 0 \pmod{p}$:

$$f(x) \equiv f'(x)h'(x) \pmod{p}, \qquad g(x) \equiv g'(x)h'(x) \pmod{p}.$$

We now make use of the following principle:

SCHATANOVSKI'S PRINCIPLE.† *If $\phi(y_1, y_2, \cdots, y_k)$ is an integral symmetric function of the indeterminates $y_1, y_2, \cdots, y_k$ with integral coefficients, and if for a natural number m*

$$f(x) \equiv (x - \alpha_1)(x - \alpha_2) \cdots (x - \alpha_k) \equiv (x - \gamma_1)(x - \gamma_2) \cdots (x - \gamma_k) \pmod{m}$$

where $f(x)$ is a polynomial with integral coefficients, then

$$\phi(\alpha_1, \alpha_2, \cdots, \alpha_k) \equiv \phi(\gamma_1, \gamma_2, \cdots, \gamma_k) \pmod{m}. \tag{7.3}$$

Let

$$\phi(y_1, y_2, \cdots, y_k) = \prod \left(\frac{y_i^n - \beta_i^n}{y_i - \beta_j} \right),$$

and let $\gamma_1, \gamma_2, \cdots, \gamma_k$ be the roots of $f'(x)h'(x) = 0$ in a definite order. Then on taking $m = p$, (7.3) gives us

$$R_n = R_n(f, g) \equiv R_n(f'h', g) \pmod{p}.$$

† See Schatanovski [1], Lubelski [1], [2]. The principle is also used constantly in Ward [2].

Here and later if $R_{f,g}$ vanishes, we can replace the congruence by an equality. A second application of Schatanovski's principle gives us

$$R_n = R_n(f, g) \equiv R_n(f'h', g'h') \equiv P^*_{k,l,m}(c', d', e') \pmod{p}.$$

Hence we obtain from (6.3) the congruence

$$(7.4) \quad R_n \equiv n^m e_m'^{n-1} P_{k-m,l-m}(c', d') P_{k-m,m}(c', e') P_{m,l-m}(e', d') Q_m^2(e') \pmod{p}.$$

In particular then $R_p \equiv 0 \pmod{p}$. Since p has no proper divisors and $R_1 = 1$, we thus obtain the following theorem:

THEOREM 7.1. *p is a rank of apparition of any prime p in (R) which divides the resultant $R_{f,g}$ of the polynomials $f(x)$ and $g(x)$ which generate (R).*

Now clearly

$$c_k \equiv c'_{k-m} e'_m \pmod{p}, \quad d_l \equiv d'_{l-m} e'_m \pmod{p}, \quad (c'_{k-m}, d'_{l-m}) \not\equiv 0 \pmod{p}.$$

Hence $e_m' \equiv 0 \pmod{p}$ if and only if $c_k \equiv d_l \equiv 0 \pmod{p}$.

Also $R_{f',g'} \not\equiv 0 \pmod{p}$, $(R_{f',h'}, R_{g',h'}) \not\equiv 0 \pmod{p}$. If we assume that $R_{f',h'} \equiv 0 \pmod{p}$, we have a congruence similar to (7.4) for $P_{k-m,m}(c', e')$; with an obvious extension of notation

$$P_{k-m,m}(c', e') \equiv n^{m'} e''^{n-1}_{m'} P_{k-m-m',m'}(c'', e'') \cdots \pmod{p}.$$

By what we have just shown, $e''_{m'} \equiv 0 \pmod{p}$ if and only if $e_{m'} \equiv c'_{k-m} \equiv 0 \pmod{p}$. A like result holds if $R_{g',h'} \equiv 0 \pmod{p}$. Now it is easily seen that in case 1, p is not a null divisor of (R). Hence we obtain the theorem:

THEOREM 7.2. *p is a null divisor of the Lucasian sequence (R) if and only if p divides the constant terms c_k and d_l of the polynomials $f(x)$ and $g(x)$ which generate (R).*

Hence if p is not such a null divisor of (R), the determination of its places of apparition in case 2 reduces by virtue of (7.4) to determining its places of apparition in various sequences dividing (R) modulo p but for which p does not divide the associated resultant. For (Ward [1] Theorem 6.3) the set of places of apparition in the product of two or more sequences is the union of the sets of places of apparition in the constituent sequences, and the ranks of apparition in the product are immediately specifiable in terms of the ranks of apparition in the constituents. *It suffices therefore to consider only case* 1.

8. We next prove that it suffices to consider the case when $f(x)$ and $g(x)$ are irreducible modulo p. With our previous notation, let p be a prime which does not divide the resultant of the generators of (R). Let the decompositions of the polynomials $f(x)$ and $g(x)$ modulo p be

$$f(x) \equiv f_1(x)^{a_1} \cdots f_r(x)^{a_r} \pmod{p}, \qquad g(x) \equiv g_1(x)^{b_1} \cdots g_s(x)^{b_s} \pmod{p}.$$

Here the polynomiials $f_1(x), \cdots, g_s(x)$ have integral coefficients, and are primary, irreducible, and co-prime in pairs modulo p. Schatanovski's principle gives us then the congruence

$$R_n \equiv R_n(f, g) \equiv R_n(f_1^{a_1} f_2^{a_2} \cdots f_r^{a_r}, g_1^{b_1} g_2^{b_2} \cdots g_s^{b_s}) \pmod{p}.$$

On using the elementary multiplicative properties of resultants (Fricke [1]) this last congruence may be written

$$R_n(f, g) \equiv \{R_n(f_1, g_1)\}^{a_1 b_1} \cdots \{R_n(f_r, g_s)\}^{a_r b_s} \pmod{p}.$$

Hence it follows as in §3 that we may confine ourselves to the case where the generators of (R) are irreducible modulo p.

9. In determining the law of apparition of primes in the sequence (U), we can similarly confine ourselves to the case when the generator of (U) is irreducible modulo p. It would at first appear as if this result were a special case of the reduction for (R), since (U) is obtainable from (R) by setting $g(x) = df(x)/dx$. But the leading coefficient of df/dx is not unity but k, so that the primes dividing k would be unclassified by this method. It is however possible to parallel the reduction for (R), and the process is so similar that we shall merely indicate the main steps.

We begin as in §6 by considering the effect upon

$$\Psi_k(y) = \prod_{\substack{i,j=1 \\ i<j}}^{k} \left(\frac{y_i^n - y_j^n}{y_i - y_j} \right) = Q_k(Y) \tag{9.1}$$

of substituting, in place of $y_1, \cdots, y_k$, h distinct new indeterminates $t_{11}, \cdots, t_{1k_1}, \cdots, t_{r1}, \cdots, t_{rk_r}$ so that we have

$$(x - y_1)(x - y_2) \cdots (x - y_k) = \prod_{u=1}^{r} \prod_{i=1}^{k_u} (x - t_{ui})^{a_u},$$

$$a_1 k_1 + a_2 k_2 + \cdots + a_r k_r = k, \qquad k_1 + k_2 + \cdots + k_r = h,$$

and at least one a_u is greater than unity. The right side of (9.1) then becomes a polynomial in the quantities $T_1, \cdots, T_r$ defined by

$$(x - t_{u1})(x - t_{u2}) \cdots (x - t_{uk_u}) = x^{k_u} - T_{u1} x^{k_u - 1} - \cdots - T_{uk_u}.$$

The value of the left side of (9.1) is then easily found to be

$$\pm n^l \prod_{u=1}^{r} T_{uk_u}^{A_u (n-1)} \prod_{u=1}^{r} \{Q_{k_u}(T_u)\}^{a_u^2} \prod_{\substack{u,v=1 \\ u<v}}^{r} \{P_{k_u, k_v}(T_u, T_v)\}^{a_u a_v}, \tag{9.2}$$

$$A_u = \tfrac{1}{2} a_u (a_u - 1), \qquad l = k_1 A_1 + k_2 A_2 + \cdots + k_r A_r$$

in analogy with formula (6.3).

Consider now the sequence (U) of §2 with the generator

$$f(x) = x^k - c_1x^{k-1} - \cdots - c_k = (x - \alpha_1)(x - \alpha_2) \cdots (x - \alpha_k)$$

and discriminant

$$D_f = \left\{ \pm \prod_{i<j} (\alpha_i - \alpha_j) \right\}^2.$$

If D_f does not vanish, we define the sequence

$$(U): \quad U_0, U_1, U_2, \cdots, \quad \text{by} \quad U_n = \Psi_k(\alpha) = Q_k(c).$$

If D_f vanishes, we have

$$f(x) = \{f_1(x)\}^{a_1}\{f_2(x)\}^{a_2} \cdots \{f_r(x)\}^{a_r}, \tag{9.3}$$

where $f_u(x) = x^{k_u} - c_{u1}x^{k_u-1} - \cdots - c_{uk_u} = (x-\tau_{u1}) \cdots (x-\tau_{uk_u})$ and $D_{f_u} \not\equiv 0$, $R_{f_u,f_v} \not\equiv 0$, $u \neq v$. We then define U_n by means of (9.2) as

$$U_n = \pm n^l \prod_{u=1}^{r} c_{uk_u}^{A_u(n-1)} \prod_{u=1}^{r} \{Q_{k_u}(c_u)\}^{a_u^2} \prod_{\substack{u,v=1 \\ u<v}}^{r} \{P_{k_u,k_v}(c_u, c_v)\}^{a_ua_v}. \tag{9.4}$$

Now consider the places of apparition of any prime p in (U). As in the case of (R), there are two cases according as p does or does not divide the discriminant D_f.

Case 1. $D_f \not\equiv 0 \pmod{p}$. Then $U_n \equiv 0 \pmod{p}$ if and only if

$$\Psi_k(\alpha) = \prod \left(\frac{\alpha_i^n - \alpha_j^n}{\alpha_i - \alpha_j} \right) \equiv 0 \pmod{p}.$$

Case 2. $D_f \equiv 0 \pmod{p}$. In this case (9.3) holds modulo p where we may assume that the polynomials $f_u(x)$ are irreducible modulo p and relatively prime in pairs modulo p. We deduce then from Schatanovski's principle that

$$U_n \equiv \pm n^l \prod_{u=1}^{r} c_{uk_u}^{A_u(n-1)} \prod_{u=1}^{r} \{Q_{k_u}(c_u)\}^{a_u^2} \prod_{\substack{u,v=1 \\ u<v}}^{r} \{P_{k_u,k_v}(c_u, c_v)\}^{a_ua_v} \pmod{p}. \tag{9.5}$$

This congruence is the analogue of (7.4). We deduce the theorems:

THEOREM 9.1. *p is a rank of apparition of any prime p in (U) which divides the discriminant D_f of the polynomial $f(x)$ which generates (U).*

THEOREM 9.2. *p is a null divisor of the Lucasian sequence (U) if and only if p divides the last two coefficients c_k and c_{k-1} of the polynomial $f(x)$ which generates (U).*

Formula (9.5) also shows us that it suffices to consider case 1 for (U) or (R). But in case 1 for (U), we have a decomposition (9.3) of $f(x)$ modulo p with all the a_u unity. Thus a decomposition (9.5) applies with all the a_u, a_v unity, all the A_u zero, and l zero. We thus deduce that it suffices in every case to assume the generators of (U) and (R) irreducible modulo p.

Formula (9.5) shows that *the law of apparition of primes in the sequence* (U) *depends on the law of apparition in* (R), for each sequence with general term $P_{k_u,k_v}(c_u, c_v)$ is a special (R) sequence.

III. Laws of apparition for R-sequences with irreducible generators

10. We shall now determine the law of apparition of primes p in (R) when the generators of (R) are irreducible modulo p.

With our previous notation, let

$$f(x) = x^k - c_1x^{k-1} - \cdots - c_k = (x-\alpha_1)(x-\alpha_2)\cdots(x-\alpha_k),$$
$$g(x) = x^l - d_1x^{l-1} - \cdots - d_l = (x-\beta_1)(x-\beta_2)\cdots(x-\beta_l)$$

be the generators of (R). Both $f(x)$ and $g(x)$ are algebraically irreducible. Let $\mathfrak{K}$ denote the galois field of the roots of $f(x)=0$ and $g(x)=0$ obtained by adjoining the $k+l$ quantities $\alpha_1, \cdots, \beta_l$ to the field of rationals.

Lemma 10.1. *p is a prime ideal of $\mathfrak{K}$.*

Proof. If C is the ring of integers of $\mathfrak{K}$, it suffices to show that the quotient ring $\mathfrak{O}/[p]$ is a field. Let $\mathfrak{R}$ as before denote the ring of rational integers, and let α be any root of $f(x)=0$, β any root of $g(x)=0$. Construct the ring $\mathfrak{o}=\mathfrak{R}[\alpha,\beta]$. Clearly $\mathfrak{O}$ contains $\mathfrak{o}$. Hence $\mathfrak{O}/[p]$ contains $\mathfrak{o}/[p]$. We shall now show that $\mathfrak{o}/[p]$ contains $\mathfrak{O}/[p]$ so that

$$\mathfrak{O}/[p] = \mathfrak{o}/[p]. \tag{10.1}$$

To prove this it suffices to show that every element of $\mathfrak{O}$ is congruent modulo p to an element of $\mathfrak{o}$. Let $\overline{D}$ be the discriminant of the field $\mathfrak{K}$. Then (Hilbert [1], Theorem 85, page 144)

$$(p, \overline{D}) = 1. \tag{10.2}$$

For since both $f(x)$ and $g(x)$ are irreducible modulo p, p is prime to their discriminants.

We can choose rational integers $e_1, \cdots, e_k$; $f_1, \cdots, f_l$ such that $\theta = e_1\alpha_1 + \cdots + e_k\alpha_k + f_1\beta_1 + \cdots + f_l\beta_l$ is a primitive element of $\mathfrak{K}$. But we have the congruences in $\mathfrak{O}$

$$\alpha_i \equiv \alpha^{p^{ri}} \pmod{p}; \quad \beta_j \equiv \beta^{p^{sj}} \pmod{p}, \quad i = 1, \cdots, k; j = 1, \cdots, l, \tag{10.3}$$

where $r_1, \cdots, r_k$; $s_1, \cdots, s_l$ are the integers $1, \cdots, k$; $1, \cdots, l$ in some order. Hence θ is congruent modulo p to an element of $\mathfrak{o}$. But if N is the degree of the field $\mathfrak{R}$ and $\overline{D}$ as before its discriminant, the N elements $\overline{D}^{-1}, \theta\overline{D}^{-1}, \cdots, \theta^{N-1}\overline{D}^{-1}$ are a basis of $\mathfrak{O}$. Hence by (10.2), each element of this basis is congruent modulo p to an element of $\mathfrak{o}$. Hence (10.1) follows.

Now the ring $\mathfrak{o}/[p]$ may be obtained either by first adjoining α and β to $\mathfrak{R}$ and then forming the quotient ring, or else by first forming the quotient ring $\mathfrak{R}/[p]$ and then adjoining α and β. Since $\mathfrak{R}/[p]$ is a field, $\mathfrak{o}/[p]$ is consequently a field, so that by (10.1), $\mathfrak{O}/[p]$ is a field.

11. Now assume that for a certain value of n

$$R_n = \prod\left(\frac{\alpha_i^n - \beta_j^n}{\alpha_i - \beta_j}\right) \equiv 0 \pmod{p}.$$

Since p is prime to the resultant of $f(x)$ and $g(x)$, we see from Lemma 10.1 that this congruence can hold if and only if

$$\alpha_i^n \equiv \beta_j^n \pmod{p} \tag{11.1}$$

in $\mathfrak{O}$ for some values of the subscripts i and j.

On multiplying (11.1) by β_j^{-n}, raising to the proper power of p, and utilizing (10.3) we obtain as a necessary and sufficient condition that p divide R_n†

$$\{\alpha\beta^{-p^s}\}^n \equiv 1 \pmod{p}, \qquad 1 \leqq s \leqq l. \tag{11.2}$$

Now $\alpha\beta^{-p^s}$ is an element of the finite galois field $\mathfrak{R}^* = \mathfrak{R}[\alpha, \beta]/[p]$ of order p^m where m is the least common multiple of the degrees of $f(x)$ and $g(x)$. Let σ_s be its period. Then (11.2) holds if and only if

$$n \equiv 0 \pmod{\sigma_s}. \tag{11.3}$$

We thus obtain the following theorem:

THEOREM 11.1. *If σ_s is the period of $\alpha\beta^{-p^s}$ modulo p in $\mathfrak{O}/[p]$, then $\sigma_1, \sigma_2, \cdots, \sigma_l$ constitute a set of generators for the multiplicative set $\mathfrak{S}_p$ of places of apparition of p in (R).*

If we regard the solution of the problem of determining the period of a mark in a finite field as known, the law of apparition of p in (R) is determined: all ranks of apparitions necessarily occur in the set $\sigma_1, \sigma_2, \cdots, \sigma_l$, and to obtain them we merely reject all σ_i which are multiples of order σ_i in the set. The rank of p is then the least common multiple of the surviving σ, and the set $\mathfrak{S}_p$ is exactly specified.

12. From a more realistic standpoint, the period of a mark in a finite

† If d_l' is chosen so that $d_l' d_l \equiv 1 \pmod{p}$, an explicit expression for β^{-1} is given by the congruence $\beta^{-1} \equiv d_l'(\beta^{l-1} - d, \beta^{l-2} - \cdots - d_{l-1}) \pmod{p}$.

field is not given to us by merely specifying the field and the mark, so that it becomes important to reduce the number of crude generators $\sigma_1, \cdots, \sigma_l$ of $\mathfrak{S}_p$ as much as possible. Before giving the details of this reduction, we shall consider the sequence (U) for the case when its generators are irreducible modulo p.

By a repetition of the arguments applied to (R) in the previous section, we deduce that if (U) is generated by a polynomial $f(x)$ which is irreducible modulo p, then

$$U_m \equiv 0 \pmod{p}$$

if and only if

$$m = 0 \pmod{\rho_s},$$

where ρ_s is the period of α^{p^s-1} in the field $\Re[\alpha]/[p]$, s is an integer $\geqq 1$ and $\leqq k$, the degree of $f(x)$, and α is any root of $f(x)=0$.

But if λ is the period of α, the period ρ_s of α^{p^s-1} is easily seen to be the residual† of p^s-1 with respect to λ. In the usual notation for residuals,

$$\rho_s = \lambda : p^s - 1. \tag{12.1}$$

We observe in particular that

$$\rho_k = 1, \qquad \rho_1 = \mu. \tag{12.2}$$

Here μ is the restricted period of $f(x)$ modulo p; that is, the least positive integer such that (Ward [2], p. 284)

$$\alpha_1^{\mu} \equiv \alpha_2^{\mu} \equiv \cdots \equiv \alpha_k^{\mu} \pmod{p}.$$

Now (Ward [4], p. 627) by (12.1)

$$\begin{aligned}[\rho_s, \rho_t] &= [\lambda : p^s - 1, \lambda : p^t - 1] = \lambda : (p^s - 1, p^t - 1) \\ &= [\lambda : p^{(s,t)} - 1]\end{aligned}$$

since the sequence $0, p-1, p^2-1, p^3-1, \cdots$ has the property that $(p^s-1, p^t-1) = p^{(s,t)}-1$ (Lucas [1], Ward [5], [6]). Thus

$$[\rho_s, \rho_t] = \rho_{(s,t)}. \tag{12.3}$$

It follows from (12.3) that if *s divides t, then ρ_t divides ρ_s*. On taking $t=k$ in (12.3) and using (12.2), we see that

$$\rho_s = \rho_d \text{ where } d = (s, k) \text{ divides } k. \tag{12.4}$$

† For the properties of residuals used here, see Ward [3], [4].

We therefore need consider only periods ρ_d where d divides k. But† if $d \mid d' \mid k$, then $\rho_{d'} \mid \rho_d$.

We therefore need consider only periods ρ_d where d divides k and no multiple of d divides k. On collecting these results, we obtain Theorem 5.1 of the introduction.

Let $\pi_{ks} = p^k - 1 : p^s - 1$. Since λ divides $p^k - 1, \lambda : p^s - 1$ divides $p^k - 1 : p^s - 1$. (Ward [4] formula (4.51)). Hence $\rho_s \mid \pi_{ks}$, $(s = 1, 2, 3, \cdots, k)$.

We thus obtain from Theorem 5.1 the following result which gives us a useful upper limit to the ranks of apparition of p.

THEOREM 12.1. *If $f(x)$ is irreducible modulo p and of degree k, the ranks of apparition of p in the sequence (U) generated by $f(x)$ divide the numbers $p^k - 1/p^{k/q_1} - 1$, $p^k - 1/p^{k/q_2} - 1, \cdots, p^k - 1/p^{k/q_K} - 1$. Here $q_1, \cdots, q_K$ are the K prime factors of k.*

If $Q = q_1 q_2 \cdots q_K$, it easily follows that *the rank of p in (U) must divide the number $p^k - 1/p^{k/Q} - 1$.*

13. We return now to the reduction of the generators of the places of apparition of p in (R). With the notation of §10, let γ be a primitive element of the finite field $\mathfrak{R}^*$. Then

$$\alpha \equiv \gamma^a, \qquad \beta \equiv \gamma^b, \qquad \alpha\beta^{-p^s} \equiv \gamma^{a - bp^s} \pmod{p}$$

where a and b are positive integers.‡ Hence

$$\sigma_s = p^m - 1 : (a - bp^s), \qquad (s = 1, 2, \cdots, l). \tag{13.1}$$

Here m, it will be recalled, is the least common multiple of the degrees of the generators of (R).

We extend the definition of the σ_s over the entire ring $\mathfrak{R}$ by letting

$$\sigma_r = \sigma_s \quad \text{if} \quad r \equiv s \pmod{l}. \tag{13.2}$$

The numbers σ_s have the following strange property which stands in remarkable contrast to the property of the ranks of apparition of p in (U) expressed by formula (12.3).

THEOREM 13.1. *Let p be a prime, let the generators of the sequence be irreducible modulo p, and let m be the least common multiple of their degrees. Then the least common multiples of pairs of generating elements for the places of apparition of p in (R) satisfy the relation*

$$[\sigma_s, \sigma_t] = [\sigma_s, \sigma_{s \pm (m, t-s)}]. \tag{13.3}$$

† We use the usual notation $a \mid b$ for a divides b.

‡ If λ_1 and λ_2 are the periods of $f(x)$ and $g(x)$ modulo p, the numbers a and b are subject to the conditions

$$(a, p^m - 1) = p^m - 1 : \lambda_1, \qquad (b, p^m - 1) = p^m - 1 : \lambda_2.$$

Proof. For convenience write $r+s$ in place of t so that (13.3) becomes

$$[\sigma_s, \sigma_{r+s}] = [\sigma_s, \sigma_{r\pm(m,r)}]. \tag{13.31}$$

By (13.1) and elementary properties of residuals

$$\begin{aligned}[\sigma_s, \sigma_{r+s}] &= p^m - 1:(p^m - 1, a - bp^s, a - bp^{r+s}),\\ [\sigma_s, \sigma_{s\pm(m,r)}] &= p^m - 1:(p^m - 1, a - bp^s, a - bp^{s\pm(m,r)}).\end{aligned} \tag{13.4}$$

Thus the proof reduces to showing that the two greatest common divisors on the right of (13.4) are equal. Now

$$\begin{aligned}(p^m - 1, a - bp^s, a - bp^{r+s}) &= (p^m - 1, p^s b(p^r - 1), a - bp^s)\\ &= (p^m - 1, b(p^r - 1), a - bp^s)\end{aligned}$$

since $(p^s, p^m-1)=1$; and we obtain

$$(p^m-1, a-bp^s, a-bp^{r+s}) = (p^m-1, b(p^{(m,r)}-1), a-bp^s)$$

since $(p^r-1, p^m-1) = p^{(m,r)}-1$. Hence since $(p^{s-(m,r)}, p^m-1)=1$ and $(p^s, p^m-1)=1$,

$$\begin{aligned}(p^m - 1, a - bp^s, a - bp^{r+s}) &= (p^m - 1, p^{s-(m,r)}b(p^{(m,r)} - 1), a - bp^s)\\ &= (p^m - 1, a - bp^{s-(m,r)}, a - bp^s),\\ (p^m - 1, a - bp^s, a - bp^{r+s}) &= (p^m - 1, p^s b(p^{(m,r)} - 1), a - bp^s)\\ &= (p^m - 1, a - bp^{s+(m,r)}, a - bp^s).\end{aligned}$$

It follows from (13.3) that the $l(l-1)/2$ least common multiples $[\sigma_s, \sigma_t]$, $(s, t=1, \cdots, l; s<t)$, may be grouped into a certain number of sets such that all the members of a set are equal to one another.* For example, if $l=6$, $k=2$, we find that the fifteen least common multiples are grouped into six sets:

$$\begin{aligned}&[\sigma_1, \sigma_2] = [\sigma_2, \sigma_3] = [\sigma_3, \sigma_4] = [\sigma_4, \sigma_5] = [\sigma_5, \sigma_6] = [\sigma_1, \sigma_6];\\ &[\sigma_1, \sigma_3] = [\sigma_1, \sigma_5] = [\sigma_3, \sigma_5]; \ [\sigma_2, \sigma_4] = [\sigma_4, \sigma_6] = [\sigma_2, \sigma_6];\\ &[\sigma_1, \sigma_4]; \quad [\sigma_2, \sigma_5]; \quad [\sigma_3, \sigma_6].\end{aligned}$$

The case when there is only one such set is of particular interest on account of the following easily proved theorem:

THEOREM 13.2. *If all of the $l(l-1)/2$ least common multiples $[\sigma_s, \sigma_t]$ are equal to one another, then if there is more than one rank of apparition of p in (R), the rank of p in (R) is the least common multiple of the two smallest σ_t. If the smallest σ_t divides the next smallest, there is only one rank of apparition.*†

* But not necessarily unequal to least common multiples in other sets.

† It must not be supposed that there are at most two ranks of apparition. For instance if $l=3$, we might conceivably have $\sigma_1=6$, $\sigma_2=10$, $\sigma_3=15$. The least common multiples $[\sigma_s, \sigma_t]$ then equal 30.

It can be shown from (13.3) by a simple enumeration that the hypothesis of the theorem is satisfied if $l=2$, 3 or 5; $l=6$ and $k\equiv 0 \pmod 4$; $l=7$ and $k\not\equiv 0 \pmod{60}$.

14. If we raise the congruence $\alpha^n \equiv \beta^{p^s n} \pmod p$ to the p^lth and p^kth powers successively, we obtain $\alpha^{(p^l-1)n}\equiv 1 \pmod p$, $\beta^{p^s(p^k-1)n}\equiv 1 \pmod p$. Hence if λ_1 and λ_2 denote the periods of $f(x)$ and $g(x)$,

$$n \equiv 0 \pmod{\lambda_1 : p^l - 1}, \qquad n \equiv 0 \pmod{\lambda_2 : p^k - 1},$$

where we are using the notation already employed in §12 for residuals. Now $\lambda_1 : p^l-1 = \lambda_1 : (\lambda_1, p^l-1) = \lambda_1 : (\lambda_1, p^k-1, p^l-1)$ since λ_1 divides p^k-1. But $(p^k-1, p^l-1) = p^{(k,l)}-1$. Hence $\lambda_1 : p^l-1 = \lambda_1 : p^{(k,l)}-1$. Similarly $\lambda_2 : p^k-1 = \lambda_2 : p^{(k,l)}-1$. Hence $n\equiv 0 \pmod{[\lambda_1 : p^{(k,l)}-1, \lambda_2 : p^{(k,l)}-1]}$ or

$$n \equiv 0 \pmod{[\lambda_1, \lambda_2] : p^{(k,l)} - 1}. \tag{14.1}$$

(14.1) gives us a *lower limit* for every rank of apparition σ of p in (R) in terms of the periods of the generators of (R). An upper limit may be obtained as follows:

If μ_1, μ_2 denote the restricted periods of $f(x)$ and $g(x)$ respectively; then

$$\alpha_1^{\mu_1} \equiv \alpha_2^{\mu_1} \equiv \cdots \equiv \alpha_k^{\mu_1} \equiv a \pmod p, \qquad \beta_1^{\mu_2} \equiv \beta_2^{\mu_2} \equiv \cdots \equiv \beta_l^{\mu_2} \equiv b \pmod p,$$

where a and b are rational integers. Then if ϕ is the least positive value of x such that $a^x \equiv b^x \pmod p$, every other such x is easily shown to be divisible by ϕ. Now ϕ as a divisor of $p-1$ is relatively prime to the restricted periods μ_1 and μ_2 (Ward [5]) and hence relatively prime to their least common multiple $[\mu_1, \mu_2]$. It readily follows that *the least positive value of n such that*

$$\alpha_1^n \equiv \alpha_2^n \equiv \cdots \equiv \alpha_k^n \equiv \beta_1^n \equiv \beta_2^n \equiv \cdots \equiv \beta_l^n \pmod p \tag{14.2}$$

is $\mu = \phi[\mu_1, \mu_2]$. *Every other such n is divisible by* μ. Since (14.2) is satisfied for $n = [\lambda_1, \lambda_2]$ we see that $\phi \mid [\lambda_1, \lambda_2]/[\mu_1, \mu_2]$.

It is now easy to show (compare M. Hall [1] or Ward [2]) that *every rank of apparition of p in* (R) *divides* μ. We thus obtain the following theorem:

THEOREM 14.1. *Let the generators of* (R) *be irreducible modulo p with degrees k and l and with periods and restricted periods* λ_1, μ_1 *and* λ_2, μ_2 *respectively. Then for every rank of apparition* σ *of p in* (R),

$$[\lambda_1, \lambda_2] : (p^{(l,k)} - 1) \tag{14.3}$$

divides σ; σ *divides* $\phi[\mu_1, \mu_2]$. *Here* ϕ *divides* $[\lambda_1, \lambda_2]/[\mu_1, \mu_2]$, *and* $\mu = \phi[\mu_1, \mu_2]$ *is the least positive value of n such that the congruence* (14.2) *holds.*

In particular if l and k are co-prime, $[\lambda_1, \lambda_2] : p^{(l,k)} - 1 = [\mu_1, \mu_2]$. Hence if σ is a rank of apparition of p so that $\alpha_i^\sigma \equiv \beta_j^\sigma \pmod p$, (14.3) implies that μ divides σ.

THEOREM 14.2. *If the generators of* (R) *are irreducible modulo* p *and if their degrees are relatively prime, there is only one rank of apparition of* p *in* (R). *This rank is the least positive value of* n *such that the congruence* (14.2) *holds, and it is a multiple of the least common multiple of the restricted periods of the generators of* (R), *and a divisor of the least common multiple of their periods.*

IV. APPLICATIONS TO GENERAL LUCASIAN SEQUENCES

15. We shall now prove Theorem 2.1 of the introduction. Let (u): $u_0, u_1, u_2, \cdots$ be a Lucasian sequence belonging to the polynomial $f(x)=x^k-\cdots-c_k=(x-\alpha_1)\cdots(x-\alpha_k)$, and let p be any prime dividing neither its constant term* c_k nor its discriminant $D=D_f=\pm\prod_{i<j}(\alpha_i-\alpha_j)^2$.

Let $\mathfrak{R}$ now denote the galois field of the roots of $f(x)=0$ and $\mathfrak{p}$ a prime ideal divisor of p in $\mathfrak{R}$. Then the general term u_n of (u) is of the form

$$u_n = A_1\alpha_1^n + \cdots + A_k\alpha_k^n,$$

where $DA_1, \cdots DA_k$ are integers of $\mathfrak{R}$, so that $A_1, \cdots, A_k$ are integers modulo $\mathfrak{p}$. Since (u) is a divisibility sequence, $u_n\equiv 0 \pmod{p}$ if and only if

$$A_1\alpha_1^{mn} + A_2\alpha_2^{mn} + \cdots + A_k\alpha_k^{mn} \equiv 0 \pmod{\mathfrak{p}}, \qquad m = 1, 2, \cdots, k.$$

Thus the determinant of this system of congruences must be divisible by $\mathfrak{p}$. This determinant may be written $c_k^n\prod_{i<j}(\alpha_i-\alpha_j)U_n$. Since $\mathfrak{p}$ is prime to the first two terms, $U_n\equiv 0 \pmod{\mathfrak{p}}$ so that $U_n\equiv 0 \pmod{p}$. Hence every place of apparition of p in (u) is also a place of apparition of p in (U).

16. Suppose that the k (not necessarily distinct) nth powers of the roots of $f(x)=0$ are grouped modulo $\mathfrak{p}$ into t incongruent sets:

$$\alpha_{i_1}^n \equiv \alpha_{i_2}^n \equiv \cdots \equiv \alpha_{i_{s_i}}^n \equiv \zeta_i \pmod{\mathfrak{p}}, \qquad i = 1, 2, \cdots, t, \tag{16.1}$$

$$\zeta_i \not\equiv \zeta_j \pmod{\mathfrak{p}} \quad \text{if} \quad i \neq j; \quad s_1 + s_2 + \cdots + s_t = k.$$

Furthermore let

$$A_i = A_{i_1} + A_{i_2} + \cdots + A_{i_{s_i}}, \qquad i = 1, 2, \cdots, t. \tag{16.2}$$

THEOREM 16.1. *Any prime* p *which does not divide the discriminant of* $f(x)$ *divides a term* u_n *of the Lucasian sequence* (u) *belonging to* $f(x)$ *if and only if*

$$A_i \equiv 0 \pmod{\mathfrak{p}}, \qquad (i = 1, 2, \cdots, t).$$

Here A_i *is given by formulas* (16.1), (16.2)† *and* $\mathfrak{p}$ *is any prime ideal divisor of* p *in the Galois field of the roots of* $f(x)=0$.

* If we are willing to assume that $u_0=0$, we may dispense with this first assumption. Marshall Hall [1] has shown that u_0 is usually zero.

† The groupings of the roots in (16.1) depend of course on our choice of $\mathfrak{p}$.

Proof. See Ward [2], pp. 284–285.

We may make this result more explicit by the use of Schatanovski's principle. Suppose that the decomposition of $f(x)$ modulo p is

$$f(x) \equiv f_1(x)f_2(x) \cdots f_r(x) \pmod{p},$$

where $f_i(x)$ is primary and irreducible modulo p and of degree k_i, and let the roots of $f_i(x)=0$ be $\gamma_1^{(i)}, \gamma_2^{(i)}, \cdots, \gamma_{k_i}^{(i)}$.

Then by Schatanovski's principle

$$u_n \equiv u_n^{(1)} + u_n^{(2)} + \cdots + u_n^{(r)} \pmod{p},$$

where

$$u_n^{(i)} = \Gamma_1^{(i)}\{\gamma_1^{(i)}\}^n + \cdots + \Gamma_{k_i}^{(i)}\{\gamma_{k_i}^{(i)}\}^n$$

satisfies the difference equation associated with $f^{(i)}(x)$ and

$$\Gamma_j^{(i)} = u(\gamma_j^{(i)})/f'(\gamma_j^{(i)})$$

(Ward [2], *p.* 283).

Construct the galois field $\mathfrak{L}=\mathfrak{R}(\gamma_1^{(1)}, \cdots, \gamma_{k_r}^{(r)})$, and let $\mathfrak{M}$ be the ring of integers of $\mathfrak{L}$. Then as in §10, p is a prime ideal of $\mathfrak{L}$, for p is prime to the discriminants and resultants of all the $f_i(x)$. Furthermore the ring $\mathfrak{L}/[p]$ is a finite field of order p^H where $H=[k_1, k_2, \cdots, k_r]$.

Suppose that in $\mathfrak{M}$ the nth powers of the roots of $f_1(x)=0, \cdots, f_r(x)=0$ are grouped modulo p into incongruent sets as in (16.1) so that we have, omitting subscripts,

$$\{\gamma^{(i)}\}^n \equiv \{\gamma^{(j)}\}^n \pmod{p}, \qquad i \neq j. \tag{16.3}$$

Then we deduce as in §14 that

$$n \equiv 0 \pmod{[\lambda^{(i)}, \lambda^{(j)}]: p^{(k_i,k_j)} - 1}. \tag{16.4}$$

Here $\lambda^{(i)}$ and $\lambda^{(j)}$ are the periods of $f^{(i)}(x)$ and $f^{(j)}(x)$ modulo p, and k_i and k_j their degrees.

In particular, if $(k_i, k_j)=1$, then $[\lambda^{(i)}, \lambda^{(j)}]: p^{(k_i,k_j)}-1 = [\mu^{(i)}, \mu^{(j)}]$, where $\mu^{(i)}$ and $\mu^{(j)}$ are the restricted periods of $f^{(i)}(x)$ and $f^{(j)}(x)$. Now $\mu^{(i)}$ divides $p^{k_i}-1/p-1$, $\mu^{(j)}$ divides $p^{k_j}-1/p-1$, and

$$\left(\frac{p^{k_i}-1}{p-1}, \frac{p^{k_j}-1}{p-1}\right) = 1.$$

Hence we obtain from (16.4) the following theorem:

THEOREM 16.2. *If the degrees of $f^{(i)}(x)$ and $f^{(j)}(x)$ are relatively prime to one another, then the congruence* (16.3) *can hold if and only if n is divisible by the product of the restricted periods of $f^{(i)}(x)$ and $f^{(j)}(x)$.*

In the simple case when $f(x)$ is irreducible modulo p, we easily find as in §12 that $u_n \equiv 0 \pmod{p}$ only if $n \equiv 0 \pmod{\lambda : p^d - 1}$.* Here d is some divisor of k and λ is the period of $f(x)$. In particular then if k is a prime number, there is only one rank of apparition of p in (u), the restricted period of (u).

It seems unprofitable to investigate the law of apparition in general Lucasian sequences in very much greater detail until it is definitely known whether or not Lucasian sequences exist which cannot be exhibited as divisors of R-sequences.

REFERENCES

R. FRICKE
1. *Algebra*, vol. 1, Braunschweig, 1926.
M. HALL
1. American Journal of Mathematics, vol. 58 (1936), pp. 577–584.
2. Journal of the London Mathematical Society, vol. 8 (1933), pp. 162–166.
D. HILBERT
1. *Die Theorie der Algebraischen Zahlkörper*.
D. H. LEHMER
1. *An extended theory of Lucas functions*, Annals of Mathematics, (2), vol. 31 (1930), pp. 419–448.
2. Annals of Mathematics, (2), vol. 34 (1933), pp. 461–472.
S. LUBELSKI
1. Journal für die Reine und Angewandte Mathematik, vol. 102 (1930), pp. 66–67.
2. Prace Matematyczno-Fizyczne, vol. 43 (1936), p. 214.
E. LUCAS
1. American Journal of Mathematics, 1878, pp. 184–239, 289–321.
2. *Théorie des Nombres*, Paris, 1891.
T. A. PIERCE
1. Annals of Mathematics, (2), vol. 18 (1916–1917), pp. 51–64.
P. POULET
1. L'Intermédiaire des Mathématiciens, vol. 27, pp. 86–87; (2), vol. 1, p. 47; vol. 3, p. 61.
SCHATANOVSKI
1. Bulletin de la Société Physico-Mathématique de Kazan, (2), vol. 12 (1902), pp. 33–49 (in Russian).
M. WARD
1. Annals of Mathematics, (2), vol. 39 (1938), pp. 210–219.
2. These Transactions, vol. 41 (1937), pp. 276–286.
3. American Journal of Mathematics, vol. 59 (1937), pp. 921–926.
4. Duke Mathematical Journal, vol. 3 (1937), pp. 627–636.
5. Bulletin of the American Mathematical Society, vol. 40 (1934), pp. 279–281.
6. Annals of Mathematics, (2), vol. 38 (1937), pp. 725–732.

* Dr. Marshall Hall has informed me by letter that he has also obtained this result.

CALIFORNIA INSTITUTE OF TECHNOLOGY,
PASADENA, CALIF.

STOCHASTIC PROCESSES WITH AN INTEGRAL-VALUED PARAMETER*

BY

J. L. DOOB

The purpose of this paper is to set up the measure relations of the most general stochastic process and to discuss the properties of the conditional probability functions of the processes depending on a parameter running through integral values. In particular, the study of temporally homogeneous processes of this type is shown to be essentially the study of measure preserving transformations. The well known results in the latter field are applied to develop and extend the theory of Markoff processes from a new point of view.

Throughout the paper, any non-negative completely additive function of point sets, defined on a Borel field of sets† of some abstract space Ω will be called a probability measure if the space Ω is itself in the field of definition and if the set function is defined as 1 on Ω.

1. **Probability measures defined on spaces of infinitely many dimensions.** Let Ω_0 (X) be any abstract space, consisting of elements ω_0 (x), and let F_{ω_0} (F_x) be a Borel field of subsets of $\Omega_0(X)$. We shall suppose that $\Omega_0 \in F_{\omega_0}$, and $X \in F_x$. If $f(\omega_0)$ is a function defined on Ω_0 which takes on values in X, and if the Ω_0-set defined by the condition $f(\omega_0) \in E$ is in F_{ω_0} for every set E of F_x, then the function $f(\omega_0)$ will be called measurable on Ω_0. Let $\{f_n(\omega_0)\}$ be any sequence of such measurable functions, where the subscript n ranges through any aggregate $\mathcal{A}$, not necessarily denumerable. If a probability measure $P_0(\Lambda_0)$ is defined on the sets Λ_0 of F_{ω_0}, the measurable function $f_n(\omega_0)$, considered from the standpoint of probability, is a chance variable $\mathfrak{x}_n$. The following method of analyzing the mutual relations of such a set of chance variables has been used, more or less explicitly, in recent years.‡ Consider the space Ω whose points are the aggregates ω: $\{x_n\}$, $n \in \mathcal{A}$, $x_n \in X$.§ If n runs through all real numbers t, Ω consists of all functions of the real variable t,

* Presented to the Society, April 9, 1937; received by the editors July 26, 1937.

† A field of sets is a collection of sets E containing, with E_1 and E_2, their sum E_1+E_2 and difference $E_1-E_1 \cdot E_2$. A Borel field of sets is a field which contains with $E_1, E_2, \cdots$ their sum $\sum_1^\infty E_j$.

‡ Cf. Doob (I); Hopf (I); Khintchine (I); Kolmogoroff (II, pp. 24–30); Lévy (I and numerous papers); Łomnicki and Ulam (I); Paley and Wiener (I, chaps. 9 and 10, and earlier papers by Wiener). The Roman numerals refer to the bibliography at the end of the paper.

§ Each subscript n determines a coordinate x_n, and the space Ω is thus a Cartesian space with a dimension corresponding to each element of $\mathcal{A}$.

with range in X; if $\mathcal{A}$ is the set of natural numbers, Ω is the space of all sequences $(x_1, x_2, \cdots)$, $x_n \in X$; if $\mathcal{A}$ is the set of all integers $\cdots, -1, 0, 1, \cdots$, Ω is the space of all sequences $(\cdots, x_{-1}, x_0, x_1, \cdots)$, $x_n \in X$. The last example will be the one studied in later sections, but in the present section no restrictions will be made on $\mathcal{A}$. Let $\alpha_1, \cdots, \alpha_p$ be any finite set of subscripts, and let $E_1, \cdots, E_p$ be sets of F_x. The conditions

$$x_{\alpha_j} \in E_j, \qquad j = 1, \cdots, p, \tag{1.1}$$

determine a set of elements of Ω. The class of all Ω-sets defined in this way determines a Borel field F_ω of sets of Ω.* Evidently $\Omega \in F_\omega$. We shall define a probability measure $P(\Lambda)$ on the sets Λ of F_ω which will have as its value, on the set defined by (1.1), the P_0-measure of the Ω_0-set determined by the conditions

$$f_{\alpha_j}(\omega_0) \in E_j, \qquad j = 1, \cdots, p. \tag{1.2}$$

The P-measure on Ω is defined by means of a mapping of Ω_0 on Ω, which takes the sets of F_{ω_0} into sets of F_ω. Let ω_0 be a point of Ω_0. The map takes ω_0 into the point (ξ_n) of Ω defined by the equalities

$$\xi_n = f_n(\omega_0), \qquad n \in \mathcal{A}. \tag{1.3}$$

To the Ω-set determined by the conditions of (1.1) will then correspond the Ω_0-set determined by the conditions of (1.2). Then to any set Λ of F_ω will correspond a set Λ_0 in the Borel field of sets determined by those sets which are defined by conditions of type (1.2). Since the latter sets are in F_{ω_0}, $\Lambda_0 \in F_{\omega_0}$. We define $P(\Lambda)$ as $P_0(\Lambda_0)$. In this way the study of the mutual measure relations of the aggregate of functions $\{f_n(\omega_0)\}$ (the probability relations of the aggregate of chance variables $\{\mathbf{x}_n\}$) is reduced to the study of the properties of the space Ω. The earlier representation of the chance variable $\mathbf{x}_m$ by means of the function $f_m(\omega_0)$ defined on Ω_0 has been replaced by a new representation by means of the function $x_m(\omega)$, defined on Ω and taking on the value ξ_m at the point ω: (x_n) for which the mth coordinate is ξ_m.†

* The (Borel) field determined by a given collection of point sets can be defined as the intersection of all the (Borel) fields of sets which include the sets of the given collection.

† The measure relations of Ω correspond to similar relations between the chance variables $\{\mathbf{x}_n\}$; that is, between the original functions $f_n(\omega_0)$ in the sense that, if Λ is an Ω-set in the field F_ω, the corresponding Ω_0-set in the field F_{ω_0} is defined by conditions on the f's which, when imposed on the $\mathbf{x}$'s define Λ; and $P(\Lambda) = P_0(\Lambda_0)$. Due to the fact that the transformation from Ω_0 to Ω is not one-to-one, certain relations of the functions $\{f_n(\omega_0)\}$ may become distorted; thus an Ω_0-set in the field F_{ω_0} may not go into an Ω-set in the field F_ω. For example, if $\mathcal{A}$ is the set of real numbers t, so that Ω becomes the space of functions $x_t = x(t)$, and if X is the space of real numbers, it may be that $f_t(\omega_0)$ is a continuous function of t for all ω_0; on the other hand the set of elements $x(t)$ of Ω which are continuous functions of t will never be measurable (in terms of P-measure). Cf. Doob (II) for a detailed treatment of this case.

As an example of the advantages of this procedure, we give a discussion of the following classical theorem:

If $\mathbf{x}_1, \cdots, \mathbf{x}_N$ *is a set of mutually independent chance variables, the expectation of their product is the product of their expectations.**

In order to treat this theorem we take X as the space of real numbers, F_x as the field of Borel sets of X (or some more inclusive field), and $\mathcal{A}$ as the set of integers $1, \cdots, N$. The space Ω becomes ordinary N-dimensional cartesian space. A probability measure is defined, in terms of the measure properties of $\mathbf{x}_j$ on the x_j-axis,† and the P-measure on Ω is determined in the usual (multiplicative) way from these separate measures.‡ The theorem in question is now an immediate consequence of Fubini's theorem on the equality of a multiple integral and the corresponding iterated integrals.§ Incidentally Fubini's theorem provides very sensitive sets of possible hypotheses: it is sufficient to suppose that the expectation of every $\mathbf{x}_j$ exists, or else that the expectation of $\mathbf{x}_1 \cdots \mathbf{x}_N$ exists.

In the preceding discussion, the given aggregate of chance variables was considered in a given representation in terms of a corresponding aggregate of measurable functions $\{f_n(\omega_0)\}$, all defined on the same space Ω_0; and a new representation was obtained in terms of the aggregate of functions $\{x_n(\omega)\}$ defined on Ω. If the chance variables are not given in some such representation, the problem becomes more difficult. A family of chance variables is generally considered as a family of entities $\{\mathbf{x}_n\}$, distinguished by a subscript n (which is usually identified in some way with the time) varying in an aggregate $\mathcal{A}$. The chance variable $\mathbf{x}_n$, which takes on values in a space X, is considered defined by a physical process in the course of its development.‖ More specifically, it is supposed that there is a Borel field F_x of sets of points

---

* If a numerically-valued chance variable $\mathbf{x}$ is represented by a measurable function f defined on a space on which some measure is defined, and if f is absolutely integrable, the expectation of $\mathbf{x}$ is defined as the integral of f. In treating this theorem we shall assume that the N chance variables are represented by N numerically-valued functions defined and measurable on a space on which some measure is defined. The fact that such a representation is always possible will appear later in this section. A recent proof of the theorem in question, with somewhat more stringent hypotheses than those to be given, and with the chance variables represented by Lebesgue measurable functions defined on the interval $0 \leqq x \leqq 1$, was published by Kac (I, pp. 47–50).

† The measure of the interval $a < x_j < b$ is defined as the probability that $a < \mathbf{x}_j < b$.

‡ Cf. Saks, *Théorie de l'Intégrale*, Warsaw, 1933, pp. 257–263, for the details for $N=2$. The P-measure is determined by the fact that the P-measure of the N-dimensional interval $a_j < x_j < b_j$, $j=1, \cdots, N$, is defined as the product of the (1-dimensional) measures of the sides $a_j < x_j < b_j$ previously defined.

§ Saks, ibid., p. 262.

‖ Thus the chance variable $\mathbf{x}_t$ may be the x-coordinate of the position of a particle (in a Brownian movement) at time t.

of X such that X is in F_x, and that if $\alpha_1, \cdots, \alpha_p$ are elements of $\mathcal{A}$, and if $E_1, \cdots, E_p$ are in F_x, a non-negative number is assigned to the p conditions

$$x_{\alpha_j} \in E_j, \qquad j = 1, \cdots, p. \tag{1.4}$$

This number is called the probability that the conditions of (1.4) are satisfied. If $\alpha_1, \cdots, \alpha_p$ are kept fixed, these probability numbers assign measures to certain sets of the space of points $(x_{\alpha_1}, \cdots, x_{\alpha_p})$, the sets being those determined by conditions of the form

$$x_{\alpha_j} \in E_j, \qquad j = 1, \cdots, p. \tag{1.4'}$$

It is supposed further that this (p-dimensional) measure function is additive for fixed subscripts $\alpha_1, \cdots, \alpha_p$, and that it can be defined on every set of the Borel field $F_{\alpha_1, \cdots, \alpha_p}$ (the field of Ω-sets determined by the sets, defined by (1.4'), on which the function is already defined), in such a way that the extended set function is a (p-dimensional) probability measure. Now consider the space Ω and field F_ω as described above. An Ω-set, determined by conditions imposed on a certain set of coordinates, will be called a cylinder set over those coordinates. It is readily shown that the field $F_{\alpha_1, \cdots, \alpha_p}$ is the field of cylinder sets of F_ω over $x_{\alpha_1}, \cdots, x_{\alpha_p}$. What was just described was therefore the determination of a probability measure on the cylinder sets of F_ω over $x_{\alpha_1}, \cdots, x_{\alpha_p}$. Moreover the various measures, obtained by varying the coordinates involved, are coherent in the sense that if Λ is a cylinder set of F_ω over $x_{\alpha_1}, \cdots, x_{\alpha_p}$ and also a cylinder set over $x_{\beta_1}, \cdots, x_{\beta_q}$, then the probability measures, assigned to Λ in the two representations, will be the same. To show that the present situation is no more general than that described above, when a probability measure was defined on all the sets of F_ω, not merely over the cylinder sets of F_ω over a finite number of coordinates, it is necessary to prove the following theorem:

THEOREM 1.1. *A set function, defined on every cylinder set of F_ω over a finite number of coordinates, which is a probability measure on the field of cylinder sets of F_ω over each such finite set of coordinates, can be so defined on the remaining sets of F_ω that it becomes a probability measure on this field.*

This theorem was proved by Kolmogoroff (I, pp. 27–30) under the additional hypothesis that X is the set of real numbers, and that F_x is the field of Borel sets of X. Daniell (I) discussed measures on Ω in the case where Ω is the set of natural numbers (with X, F_x defined as in Kolmogoroff's case) so that Ω is the space of sequences ω: $(x_1, x_2, \cdots)$. This latter case, for which in addition the chance variables concerned are independent, has been discussed by many other writers who map Ω on a linear interval and define the

measure of an Ω-set by means of the Lebesgue measure of the corresponding set on the interval.*

Let F be the collection of Ω-sets each of which is determined by conditions of the form (1.4′) or is a finite sum of such sets. Then F is a field,† and the given set function $P(\Lambda)$ is defined on the sets of F. Let $\Lambda_0, \Lambda_1, \cdots$ be sets of F. Then if $\Lambda_1, \Lambda_2, \cdots$ are disjunct, and if, in addition,

$$\Lambda_0 = \sum_1^\infty \Lambda_m, \tag{1.5}$$

we shall show that

$$P(\Lambda_0) = \sum_1^\infty P(\Lambda_m). \tag{1.6}$$

We prove (1.6) by reducing it to the corresponding result in the special case considered by Kolmogoroff. Fix a value of n, $n=\nu$, and consider the sets $E_1^{(\nu)}, E_2^{(\nu)}, \cdots$ of X which are involved in the restrictions on x_ν used to define $\Lambda_0, \Lambda_1, \cdots$.‡ We shall define a numerically-valued function $f_\nu(x)$, with domain X, so that each set $E_j^{(\nu)}$ is determined by simple inequalities imposed on $f_\nu(x)$. In order to define the function $f_\nu(x)$ we shall need the lemma which follows:

LEMMA 1.1. *Let $\mathcal{E}_1, \mathcal{E}_2, \cdots$ be any point sets of an abstract space. There is a collection of sets $\{\mathcal{E}_r'\}$ (where r is rational and $0<r<1$) with the following properties:*

(i) *each $\mathcal{E}_j$ is in the field determined by the sets $\mathcal{E}_r'$ and conversely;*
(ii) *if $r_1<r_2$, $\mathcal{E}_{r_1} \subseteq \mathcal{E}_{r_2}$;*
(iii) *if $r_m \to r$, where $r_1>r_2>\cdots$, then $\prod_1^\infty \mathcal{E}_{r_m}' = \mathcal{E}_r'$;*
(iv) *if $r_m \to 0$ $(r_m \to 1)$, then $\prod_1^\infty \mathcal{E}_{r_m}' = 0$ $(\sum_1^\infty \mathcal{E}_{r_m}' = \sum_1^\infty \mathcal{E}_j)$.*

This lemma can be proved by a modification of the proof of a similar result due to von Neumann.§ We do not suppose that there are necessarily infinitely many distinct sets $\mathcal{E}_j$. In the contrary case, there will be only a finite number of distinct sets $\mathcal{E}_r'$.

Using this lemma, we define $f_\nu(x)$ as follows: Identify the sets $\{E_j^{(\nu)}\}$ (ν fixed) with the sets $\{\mathcal{E}_j\}$ of the lemma, and set

$$f_\nu(x) = \limsup r, \qquad x \notin \mathcal{E}_r' \tag{1.7}$$

* The details of such a map can be found in Paley and Wiener (I, pp. 145–146). Łomnicki and Ulam (I) treat the case for which $\mathcal{A}$ is the set of natural numbers, and the chance variables concerned are independent (with no restriction on X); but their proof of the theorem stated above (Theorem 1.1) is defective. (The mistake is in the proof of Lemma 4, pp. 254–255.)

† Cf. Kolmogoroff (II, pp. 25–26).

‡ Except for a denumerable set of superscripts ν, there will be only one set $E_j^{(\nu)}$; namely X itself.

§ Annals of Mathematics, vol. 33 (1932), p. 602.

The set $\mathcal{E}_r'$ is characterized by the inequality

$$f_\nu(x) \leqq r. \tag{1.8}$$

Since every set $\mathcal{E}_j$ is in the field determined by the sets $\mathcal{E}_r'$, *every set* $E_j^{(\nu)}$ *is characterized by a finite number of inequalities imposed on* $f_\nu(x)$.* *Moreover, if* $\tilde{E}$ *is any Borel set of real numbers, the x-set* E *determined by the condition* $f_\nu(x) \in \tilde{E}$ *is in the field* F_x. The latter fact is apparent if E is an interval, and its truth then follows for E any Borel set.

Now consider the space $\tilde{\Omega}$ of points $\tilde{\omega}$: $\{\tilde{x}_n\}$, $n \in \mathcal{A}$, where $\tilde{x}_n$ is any real number.† Let $F_{\tilde{x}}$ be the field of Borel sets of the $\tilde{x}$-axis, and let $F_{\tilde{\omega}}$ be the Borel field of $\tilde{\Omega}$-sets determined by $F_{\tilde{x}}$ in the same way that F_ω is determined by F_x. We map Ω on $\tilde{\Omega}$, sending the point (x_n) of Ω into the point $(\tilde{x}_n)$ of $\tilde{\Omega}$ for which

$$\tilde{x}_m = f_m(x_m), \qquad m \in \mathcal{A}. \tag{1.9}$$

This mapping is a single-valued transformation of Ω into some subset of $\tilde{\Omega}$. If $\tilde{\Lambda}$ is the $\tilde{\Omega}$-set determined by the conditions

$$\tilde{x}_{\alpha_j} \in \tilde{E}_j, \qquad j = 1, \cdots, p, \tag{1.10}$$

where $\tilde{E}_1, \cdots, \tilde{E}_p$ are Borel sets, the corresponding Ω-set Λ is determined by the conditions

$$f_{\alpha_j}(x_j) \in \tilde{E}_j, \qquad j = 1, \cdots, p. \tag{1.11}$$

It then follows from the definitions of F_ω and $F_{\tilde{\omega}}$ that the Ω-sets going into cylinder sets of $F_{\tilde{\omega}}$ over $\tilde{x}_{\alpha_1}, \cdots, \tilde{x}_{\alpha_p}$ are cylinder sets of F_ω over $x_{\alpha_1}, \cdots, x_{\alpha_p}$, and that the Ω-sets going into the sets of F are sets of F_ω. We now define a set function $\tilde{P}(\tilde{\Lambda})$, on the sets $\tilde{\Lambda}$ of $F_{\tilde{\omega}}$ which are cylinder sets over a finite number of coordinates, by setting $\tilde{P}(\tilde{\Lambda}) = P(\Lambda)$, where Λ is the Ω-set containing every element ω which is taken by the transformation into an element $\tilde{\omega}$ of $\tilde{\Lambda}$. The set function $\tilde{P}(\tilde{\Lambda})$ is uniquely defined and is a probability measure on the field of cylinder sets over any fixed finite set of coordinates. According to the definition of $\tilde{\Omega}$, there are sets $\tilde{\Lambda}_0, \tilde{\Lambda}_1, \cdots$ to which correspond the sets $\Lambda_0, \Lambda_1, \cdots$ of (1.5). The sets $\tilde{\Lambda}_1, \tilde{\Lambda}_2, \cdots$ are disjunct, and

$$\tilde{\Lambda}_0 = \sum_1^\infty \tilde{\Lambda}_m. \tag{1.5'}$$

Now we are assuming Kolmogoroff's result that Theorem 1.1 is true if X is

* One can readily determine these inequalities explicitly, using the fact that any set in the field determined by the sets $\{\mathcal{E}_r'\}$ can be written in the form $E_{r_0} + (E_{r_2} - E_{r_1}) + \cdots + (E_{r_N} - E_{r_{N-1}})$, with $r_0 < r_1 \leqq r_2 \leqq \cdots \leqq r_N$, or else in the same form without the first set E_{r_0}.

† The space $\tilde{\Omega}$ is the space of numerically-valued functions with domain $\mathcal{A}$.

the space of real numbers, and if F_x is the field of Borel sets. Then $\tilde{P}(\tilde{\Lambda})$ must certainly be completely additive on its domain of definition, so that

$$\tilde{P}(\tilde{\Lambda}_0) = \sum_1^\infty \tilde{P}(\tilde{\Lambda}_m); \tag{1.6'}$$

and this is equivalent to (1.6), since $P(\Lambda_m) = \tilde{P}(\tilde{\Lambda}_m)$, $m \geqq 0$.

The proof, that the domain of definition of $P(\Lambda)$ can be extended as described in Theorem 1.1, is now immediate. By hypothesis, $P(\Lambda)$ is an additive function of sets on the field F,* and the result just proved shows that $P(\Lambda)$ is completely additive on this field. It is then possible, according to a well known extension theorem,† to extend the definition of $P(\Lambda)$ to all the sets of the Borel field determined by the sets of F (in this case the Borel field will be F_ω), in such a way that $P(\Lambda)$ becomes a completely additive function of sets on the larger field. The set function thus obtained is the probability measure described in Theorem 1.1.

2. **Definition of a stochastic process.** We can now state the definition of a stochastic process (of the type to be studied in this paper) suggested by the preceding section. Let X be any abstract space of elements x, and let Ω be the space whose elements ω are sequences $(\cdots, x_{-1}, x_0, x_1, \cdots)$ of points of X. Let F_x be a Borel field of sets of points of X, including X itself, and suppose that $E_1, \cdots, E_p$ are sets in F_x. The conditions

$$x_{\alpha_j} \in E_j, \qquad j = 1, \cdots, p, \tag{2.1}$$

($\alpha_1, \cdots, \alpha_p$ any p distinct integers) determine a cylinder set over $x_{\alpha_1}, \cdots, x_{\alpha_p}$. The class of all such cylinder sets determines the Borel field F_ω of Ω-sets.

Let $P(\Lambda)$ be a probability measure defined on the sets of F_ω. For a fixed set of coordinates $x_{\alpha_1}, \cdots, x_{\alpha_p}$, $P(\Lambda)$ becomes a probability measure defined essentially on the p-dimensional space of elements $(x_{\alpha_1}, \cdots, x_{\alpha_p})$, and the converse (Theorem 1.1) is also true. A stochastic process depending on the parameter n running through integral values is the combination of the space Ω together with a probability measure defined on the sets of a field F_ω. More precisely, the process is the changing real entity of which the above is the mathematical abstraction. Examples of stochastic processes are given in §5.

The function $x_j(\omega)$ taking on the value x_j at the point ω: $(\cdots, x_j, \cdots)$ is a measurable function on Ω‡ taking on values in X; and the sequence of functions $\cdots, x_{-1}(\omega), x_0(\omega), x_1(\omega), \cdots$ can then be considered as a representation of a sequence of chance variables $\cdots, \mathrm{x}_{-1}, \mathrm{x}_0, \mathrm{x}_1, \cdots$. Conversely,

* The field F was defined at the beginning of the proof.

† Cf. H. Hahn, Annali delle Università Toscane, Pisa, (2), vol. 2 (1933), pp. 433–436.

‡ Cf. §1.

we have seen in §1 that any such sequence of chance variables can be represented in this way.* It is sometimes useful to consider the sequence of chance variables $x_1, x_2, \cdots$. To do this we need only restrict our attention to the cylinder sets of F_ω over $x_1, x_2, \cdots$.

We shall suppose throughout this paper that the probability measure $P(\Lambda)$ is so extended that it is defined on every set Λ_1 differing from a set Λ_0 of F_ω by at most a subset of a set on which $P(\Lambda)$ vanishes, if we set $P(\Lambda_1)=P(\Lambda_0)$. The sets of this extended domain of definition will be called P-measurable. The subsets of a set of P-measure 0 are measurable and of P-measure 0. The P-measure of a P-measurable set is the greatest lower bound of the P-measures of sets $M \supseteq \Lambda$ which are finite or denumerably infinite sums of sets of the field F defined above.† It follows from this fact that if Λ is P-measurable, to every positive number ϵ corresponds a set Λ_ϵ which is a cylinder set of F_ω over a finite number of coordinates, and which has the property that $P(\Lambda \cdot C\Lambda_\epsilon) + P(C\Lambda \cdot \Lambda_\epsilon) < \epsilon$.‡ If Λ is a P-measurable cylinder set over $x_{a_1}, \cdots, x_{a_p}$, it is determined by a condition of the form $(x_{a_1}, \cdots, x_{a_p}) \in E$, where E is a p-dimensional set of points $(x_{a_1}, \cdots, x_{a_p})$. The set E will be called an $(x_{a_1}, \cdots, x_{a_p})$-set of P-measure $P(\Lambda)$.

Let $f(\omega)$ be a numerically-valued function of ω. If for every number k the inequality $f(\omega) > k$ defines a set of F_ω (a P-measurable set), $f(\omega)$ will be called measurable with respect to F_ω (P-measurable). If $f(\omega)$ is measurable with respect to F_ω, then it is P-measurable; and conversely if $f(\omega)$ is P-measurable, there is a function $f_1(\omega)$, measurable with respect to F_ω and equal to $f(\omega)$ except possibly on a set of P-measure 0.§ This can be deduced from the following fact (which in turn follows readily from the approximation of P-measurable sets by means of cylinder sets of F_ω over a finite number of coordinates) that if $f(\omega)$ is any P-measurable function, to every positive number ϵ corresponds a function $f_\epsilon(\omega)$, measurable with respect to F_ω, depending on only a finite number of coordinates, and having the property that $|f(\omega)-f_\epsilon(\omega)| \leqq \epsilon$ except perhaps on an Ω-set of P-measure $\leqq \epsilon$. Throughout the above if the given function depends only on some given set of coordinates, the approximating functions can be supposed to depend only on the same coordinates. If $f(\omega)$ is measurable with respect to F_ω, and if $\{n_j\}$ is any set of integers, $f(\omega)$ becomes a function of the coordinates $x_{n_1}, x_{n_2}, \cdots$ only, if the other coordinates are

* Cf., however, the note on p. 88 in accordance with which it may sometimes be necessary to define a stochastic process using a subspace of Ω rather than Ω itself, as in Doob (II).

† The extension theorem used in the proof of Theorem 1.1 defines $P(\Lambda)$ in precisely this way.

‡ The complement of any set Λ will be denoted by $C\Lambda$ throughout this paper.

§ The corresponding theorems for Borel and Lebesgue measurable functions are discussed by de la Vallée Poussin in his book *Intégrales de Lebesgue, Fonctions d'Ensemble, Classes de Baire*, 2d edition, Paris, 1934, pp. 34–40.

held fast, and this function of $x_{n_1}, x_{n_2}, \cdots$ will be also measurable with respect to F_ω.

In later sections we shall use the fact that if F_x is the Borel field determined by some denumerable collection of its sets, the same will be true of F_ω. Even without this hypothesis, it can be shown (by transfinite induction) that any given set Λ, in the field F_ω, is in the field F_ω' corresponding to some Borel field $F_x' \subseteq F_x$ such that F_x' is the Borel field determined by some denumerable collection of its sets. It then follows that if $f(\omega)$ is a function measurable with respect to F_ω, a subfield F_x' of F_x can be found which satisfies the denumerability condition just described and is such that $f(\omega)$ is measurable with respect to the corresponding field F_ω'.

The integral of the P-measurable function $f(\omega)$ over a P-measurable set Λ will be noted by

$$\int_\Lambda f(\omega)dP;$$

and if the domain of integration is not stated explicitly, it will be understood to be Ω. If $f(\omega)$ depends only on a finite number of coordinates $x_{\alpha_1}, \cdots, x_{\alpha_p}$, and if Λ is a cylinder set over those coordinates, we shall use the notation

$$\int_\Lambda f(\omega)P(de_{\alpha_1,\cdots,\alpha_p})$$

for the integral of $f(\omega)$ over Λ. Corresponding notation will be used for integration with respect to other probability measures.

Let $T\omega$ be the transformation taking ω: $(\cdots, x_{-1}, x_0, x_1, \cdots)$ into ω': $(\cdots, x_{-2}, x_{-1}, x_0, \cdots)$, that is, the transformation defined by

$$x_j' = x_{j-1}, \qquad j = 0, \pm 1, \pm 2, \cdots.$$

If $T\omega$ is measure preserving, the process is called temporally homogeneous.

3. **The conditional probability functions.*** Let Λ be a P-measurable set. The conditional probability function $P_{\alpha_1,\cdots,\alpha_p}(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda)$ is defined as follows. If the set M is allowed to range through the P-measurable cylinder sets over $x_{\alpha_1}, \cdots, x_{\alpha_p}$, $P(\Lambda M)$ is a non-negative completely additive function of sets M which vanishes whenever $P(M) = 0$. There is therefore a function $P_{\alpha_1,\cdots,\alpha_p}(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda)$, a non-negative P-measurable function depending only on the coordinates $x_{\alpha_1}, \cdots, x_{\alpha_p}$, satisfying

* The ideas in this and the following section are not new, but there seems to be no systematic presentation of many of them in the literature, and some of the theorems have not been previously stated explicitly. (Known theorems and definitions are stated for later reference.) The importance and usefulness of the conditional probability and conditional expectation functions have been stressed most by P. Lévy.

$$(3.1) \qquad \int_M P_{\alpha_1,\cdots,\alpha_p}(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda)dP = P(\Lambda M)$$

for every set M.* The function $P_{\alpha_1,\cdots,\alpha_p}(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda)$ is determined uniquely up to an $(x_{\alpha_1}, \cdots, x_{\alpha_p})$-set of P-measure 0. For a given set Λ, $P_{\alpha_1,\cdots,\alpha_p}(x^0_{\alpha_1}, \cdots, x^0_{\alpha_p}; \Lambda)$ is called the conditional probability of Λ if $x_{\alpha_j} = x^0_{\alpha_j}, j = 1, \cdots, p$. The subscripts determining the function are given by the subscripts in the argument, so there will be no danger of confusion if $P_{\alpha_1,\cdots,\alpha_p}(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda)$ is written simply as $P(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda)$. We shall need the following properties of $P(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda)$ which are easily derived from the definition:†

(i) If $P(\Lambda) = 1$, then $P(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda) = 1$, except possibly on an $(x_{\alpha_1}, \cdots, x_{\alpha_p})$-set of P-measure 0. If $P(\Lambda) = 0$, then $P(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda) = 0$, except possibly on an $(x_{\alpha_1}, \cdots, x_{\alpha_p})$-set of P-measure 0.

(ii) If $\Lambda_1, \Lambda_2, \cdots$ are disjunct P-measurable sets, and if $\Lambda = \sum_1^\infty \Lambda_m$,

$$P(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda) = \sum_1^\infty P(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda_m)$$

except possibly for an $(x_{\alpha_1}, \cdots, x_{\alpha_p})$-set of P-measure 0.

This implies the following fact:

(iii) If Λ', Λ'' are P-measurable, and if $\Lambda' \subseteq \Lambda''$, then

$$P(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda') \leqq P(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda'')$$

except possibly for an $(x_{\alpha_1}, \cdots, x_{\alpha_p})$-set of P-measure 0.

By taking complements in (ii) we obtain the following property:

(iv) If $\Lambda_1, \Lambda_2, \cdots$ are P-measurable sets, and if

$$\Lambda_1 \supseteq \Lambda_2 \supseteq \cdots, \qquad \prod_1^\infty \Lambda_m = \Lambda,$$

then

$$\lim_{m\to\infty} P(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda_m) = P(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda),$$

except possibly for an $(x_{\alpha_1}, \cdots, x_{\alpha_p})$-set of P-measure 0.

THEOREM 3.1. *If F_1 is any Borel field of P-measurable sets, including the set Ω, determined by a denumerable subcollection $\Lambda_1, \Lambda_2, \cdots$, and if $\alpha_1, \cdots, \alpha_p$ are any given subscripts, then $P(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda)$ can be so defined that there is an $(x_{\alpha_1}, \cdots, x_{\alpha_p})$-set E of P-measure 0 such that for $(x_{\alpha_1}, \cdots, x_{\alpha_p})$ fixed, not in E, $P(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda)$, $(\Lambda \in F_1)$, is a probability measure.*

* This definition is due to Kolmogoroff (I, pp. 41–44).

† Cf. Kolmogoroff (II, pp. 43–44).

If we identify the sets $\Lambda_1, \Lambda_2, \cdots$ with the sets $\mathcal{E}_1, \mathcal{E}_2, \cdots$ of Lemma 1.1, we obtain sets Λ_r' corresponding to the sets $\mathcal{E}_r'$ of that lemma. We then map Ω on the t-interval $0<t<1$ by the transformation which takes a point ω into t if

$$t = \text{L. U.B.}_{\omega \in \Lambda_r'} r.$$

According to properties (i), (iii), (iv) of the conditional probability functions, if r, s are rational, with $r<s$, there is an $(x_{\alpha_1}, \cdots, x_{\alpha_p})$-set E_{rs} of P-measure 0 such that if $(x_{\alpha_1}, \cdots, x_{\alpha_p}) \notin E_{rs}$,

$$P(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda_{r'}') \leqq P(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda_s'), \tag{3.2}$$

and there is an $(x_{\alpha_1}, \cdots, x_{\alpha_p})$-set E_r of P-measure 0 such that if r' approaches r from above (r, r' rational, $r \geqq 0$), and if $(x_{\alpha_1}, \cdots, x_{\alpha_p}) \notin E_r'$, then

$$\lim_{r' \to r} P(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda'_{r'}) = P(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda_r') \tag{3.3}$$

(where if $r=0$ the right side is replaced by 0). Let E' be defined by

$$E' = \sum_{r,s} E_{rs} + \sum_r E_r,$$

and suppose that $(x_{\alpha_1}, \cdots, x_{\alpha_p})$ is fixed, not in E'. Then

$$F(r) = P(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda_r')$$

is a monotone non-decreasing function of r, defined for rational values of r between 0 and 1 and continuous on the right at these values. Define $F(t)$ for every value of t in the interval $0 \leqq t < 1$ as $\lim_{r \to t} F(r)$ (r rational, $r > t$). This is consistent with the previous definition if t is rational, and $F(r)$ thus becomes a monotone non-decreasing function $F(t)$ defined for $0 \leqq t < 1$ and continuous on the right. There is a non-negative completely additive function of Borel sets on the t-interval (0, 1) determined by the condition that its value on the interval $0 < t \leqq r$ is $F(r)$.* The Borel t-sets correspond to the sets of a certain Borel field $\tilde{F}$ of Ω-sets, under the transformation from ω to t, and a non-negative completely additive set function $\tilde{P}(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda)$ is thereby defined on these Ω-sets. The Borel sets of the interval (0, 1) are the sets of the Borel field determined by intervals of the type $0 < t \leqq r$, for r rational, so that the sets of $\tilde{F}$ are the sets of the Borel field determined by the images of such intervals. The image of the interval $0 < t \leqq r$ (r rational) is the set Λ_r', so that the field $\tilde{F}$ includes every set Λ_r' and therefore every set Λ_r; $\tilde{F} \supseteq F_1$. By definition of $\tilde{P}(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda)$, if r is rational,

* J. Radon, Sitzungsberichte der Akademie der Wissenschaften, Vienna, class IIa, vol. 122 (1913), pp. 1305–1313.

$$\tilde{P}(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda_r') = P(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda_r'). \tag{3.4}$$

We deduce from this, using the properties of the conditional probability functions listed above, that if Λ is a set in the field F_1,

$$\tilde{P}(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda) = P(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda), \tag{3.5}$$

except perhaps on an $(x_{\alpha_1}, \cdots, x_{\alpha_p})$-set of P-measure 0; and to define $P(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda)$ as required in the statement of the theorem, we need only re-define $P(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda)$ as $\tilde{P}(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda)$.

If $\phi(x_{\alpha_1}, \cdots, x_{\alpha_p}, x_{\beta_1}, \cdots, x_{\beta_q}) = \phi((x_\alpha), (x_\beta))$ is a P-measurable and integrable function depending only on the indicated coordinates, we now define the function

$$E(x_{\alpha_1}, \cdots, x_{\alpha_p}; \phi) = \int \phi((x_\alpha), (x_\beta)) P(x_{\alpha_1}, \cdots, x_{\alpha_p}; de_{\beta_1, \cdots, \beta_q}) \tag{3.6}$$

not as an integral, but in such a way that the indicated integration actually gives $E(x_{\alpha_1}, \cdots, x_{\alpha_p}; \phi)$ when it can be carried out. Let M_a be a P-measurable cylinder set over $x_{\alpha_1}, \cdots, x_{\alpha_p}$. Then $E(x_{\alpha_1}, \cdots, x_{\alpha_p}; \phi)$ is defined as the P-measurable integrable function which satisfies

$$\int_{\mathrm{M}_a} E(x_{\alpha_1}, \cdots, x_{\alpha_p}; \phi) dP = \int_{\mathrm{M}_a} \phi((x_\alpha), (x_\beta)) dP \tag{3.7}$$

for all sets M_a. The function $E(x_{\alpha_1}, \cdots, x_{\alpha_p}; \phi)$ is known as the conditional expectation of ϕ for given $(x_{\alpha_1}, \cdots, x_{\alpha_p})$.* Changing ϕ on a set of P-measure 0 does not affect $E(x_{\alpha_1}, \cdots, x_{\alpha_p}; \phi)$. If ϕ is the characteristic function of a P-measurable cylinder set over $x_{\beta_1}, \cdots, x_{\beta_q}$,

$$E(x_{\alpha_1}, \cdots, x_{\alpha_p}; \phi) = P(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda).$$

THEOREM 3.2. *Suppose that $P(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda)$ can be defined so that for each $(x_{\alpha_1}, \cdots, x_{\alpha_p})$ not in some $(x_{\alpha_1}, \cdots, x_{\alpha_p})$-set of P-measure 0, $P(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda)$ becomes a probability measure for Λ in the field of cylinder sets of F_ω over $(x_{\beta_1}, \cdots, x_{\beta_q})$. Then* (3.6) *can be interpreted as ordinary integration.*

This means, if $\phi((x_\alpha), (x_\beta))$ is P-measurable and integrable, that (a) whenever $(x_{\alpha_1}, \cdots, x_{\alpha_p})$ is not in some exceptional set of P-measure 0, $\phi((x_\alpha), (x_\beta))$ for $(x_{\alpha_1}, \cdots, x_{\alpha_p}) = (x_{\alpha_1}^0, \cdots, x_{\alpha_p}^0)$ fixed is measurable in terms of the measure function $P(x_{\alpha_1}, \cdots, x_{\alpha_p}; de_{\beta_1, \cdots, \beta_q})$ (that is, that the cylinder set over $(x_{\beta_1}, \cdots, x_{\beta_q})$ defined by $\phi > k$ with $(x_{\alpha_1}, \cdots, x_{\alpha_p}) = (x_{\alpha_1}^0, \cdots, x_{\alpha_p}^0)$ is either in F_ω or differs from such a set by a subset of such a set on which $P(x_{\alpha_1}^0, \cdots,$

* This definition is due to Kolmogoroff (II, pp. 46–47).

$x^0_{\alpha_p}; \Lambda$) vanishes); and that (b) the integral (3.6) exists and is $E(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda)$, if we neglect sets of P-measure 0.

We shall suppose that $p = q = 1$ to simplify the notation, and we can then drop the subscript 1 from α and β. We shall suppose that $P(x_\alpha; \Lambda_\beta)$ is already defined to satisfy the conditions of the theorem. In order to avoid confusion we shall reserve the integral sign throughout this proof for actual integration.

(i) According to a theorem of Kolmogoroff (II, pp. 48–49), if $\phi(x_\alpha, x_\beta)$ is P-measurable and integrable, then

$$(3.8) \quad E(x_\alpha; \phi) = \lim_{\lambda \to 0} \sum_{k=-\infty}^{+\infty} k\lambda P(x_\alpha; \lambda k \leqq \phi < \lambda(k+1)), \qquad \lambda > 0,^*$$

(if λ approaches 0 taking on only a denumerable set of values), except perhaps for an x_α-set of P-measure 0. The series in (3.8) converges absolutely for each value of λ, except perhaps for an x_α-set of P-measure 0. In particular, suppose that $\phi(x_\alpha, x_\beta)$ is measurable with respect to F_ω. Then for fixed x_α, $\phi(x_\alpha, x_\beta)$ becomes a function of x_β which is measurable with respect to F_ω (cf. §2). The existence of the limit on the right, for a fixed value of x_α, is exactly a condition that the integral

$$\int \phi(x_\alpha, x_\beta) P(x_\alpha; de_\beta)$$

exist; and in fact the limit is equal to this integral. Theorem 3.2 thus follows from Kolmogoroff's result for a function which is measurable with respect to F_ω.

(ii) Let $\phi_0(x_\alpha, x_\beta)$ be the characteristic function of a set Λ_0 in the field F_ω, of P-measure 0, and let $\Lambda(\xi)$, $(\xi \in X)$, be the cylinder set over x_β determined by the condition $\phi_0(\xi, x_\beta) = 1$. Then, neglecting x_α-sets of P-measure 0 and using (i), we obtain

$$0 = E(x_\alpha, \phi_0) = \int \phi_0(x_\alpha, x_\beta) P(x_\alpha; de_\beta) = P(x_\alpha; \Lambda(x_\alpha));$$

that is, $P(x_\alpha; \Lambda(x_\alpha)) \equiv 0$. This same result will be true even if Λ_0 is only supposed P-measurable, since it can then be enclosed in a set Λ_0', in the field F_ω, which is of P-measure 0.† From this it follows that if $\phi(x_\alpha, x_\beta)$ is any function which vanishes except perhaps on an (x_α, x_β)-set of P-measure 0, then

* The function $P(x_\alpha; \lambda k \leqq \phi < \lambda(k+1))$ is the conditional probability, for given x_α, that $\lambda k \leqq \phi < \lambda(k+1)$. Kolmogoroff's result does not require the hypotheses of Theorem 3.2.

† Cf. §2. The result is a generalization of the fact that if E is a Lebesgue measurable set of measure 0, in two-dimensional (x, y) space, the intersection of E with the line $x = c$ will be of (one-dimensional) Lebesgue measure 0 for almost all values of c.

$$E(x_\alpha, \phi) = \int \phi(x_\alpha, x_\beta) P(x_\alpha; de_\beta) = 0,$$

if we neglect an x_α-set of P-measure 0.

(iii) Let $\phi(x_\alpha, x_\beta)$ be any P-measurable and integrable function depending only on x_α, x_β. Then it can be expressed in the form

$$\phi(x_\alpha, x_\beta) = \phi_0(x_\alpha, x_\beta) + \phi_1(x_\alpha, x_\beta)$$

(cf. §2), where $\phi_0(x_\alpha, x_\beta)$ vanishes except perhaps on an (x_α, x_β)-set of P-measure 0, and where $\phi_1(x_\alpha, x_\beta)$ is measurable with respect to F_ω. Combining this fact with the results of (i) and (ii) we see that (neglecting x_α-sets of P-measure 0), for fixed $x_\alpha = x_\alpha^0$, $\phi(x_\alpha^0, x_\beta)$ is measurable in x_β with respect to the measure function $P(x_\alpha^0; de_\beta)$, and that

$$\int \phi(x_\alpha, x_\beta) P(x_\alpha, de_\beta) = \int \phi_1(x_\alpha, x_\beta) P(x_\alpha; de_\beta) = E(x_\alpha; \phi_1) = E(x_\alpha; \phi),$$

as was to be proved. Conversely it is readily seen that *if* (3.6) *exists as an integral, except possibly for an* (x_α)*-set of* P*-measure* 0, *and if the function of* $(x_{\alpha_1}, \cdots, x_{\alpha_p})$ *thus obtained is integrable, then* $\phi((x_\alpha), (x_\beta))$ *is itself integrable, and the original definition of conditional expectation is applicable.*

According to Theorem 3.1, the hypotheses of Theorem 3.2 are always satisfied if F_x (and therefore F_ω) is the Borel field determined by a denumerable collection of its sets. This will be true, for example, if X is euclidean space of $N \geqq 1$ dimensions, and if F_x is the field of Borel sets of X. A stronger statement can be made, however, since if ϕ_1 is measurable with respect to F_ω, there is always (cf. §2) a Borel field $F_x(\phi_1)$ of X-sets, (depending on ϕ_1), which is the Borel field determined by a denumerable collection of its sets, such that if $F_\omega(\phi_1)$ is the Borel field of Ω-sets defined in terms of $F_x(\phi_1)$, as F_ω is defined in terms of F_x, then ϕ_1 is measurable with respect to $F_\omega(\phi_1)$. Theorem 3.1 then shows, since only sets of $F_x(\phi_1)$ (or $F_\omega(\phi_1)$) are involved, that (3.6) can always be interpreted as integration,* if we define $P(x_\alpha; \Lambda_\beta)$ in a way depending on the function ϕ under consideration.

The following theorem, proved by Kolmogoroff (II, pp. 47–48), is stated for future reference:

THEOREM 3.3. *If* α, β, γ *are distinct integers, and if* $\phi(x_\alpha, x_\beta, x_\gamma)$ *is* P*-measurable, then*

$$E(x_\alpha; \phi) = E[x_\alpha; E(x_\alpha, x_\beta; \phi)]. \tag{3.9}$$

* The transition from the P-measurable function ϕ to the function ϕ_1, measurable with respect to F_ω, is made as in (ii) above.

In particular, if ϕ is the characteristic function of a P-measurable cylinder set Λ, over x_γ,

$$(3.10)\qquad P(x_\alpha; \Lambda) = E[x_\alpha; P(x_\alpha, x_\beta; \Lambda)] = \int P(x_\alpha, x_\beta; \Lambda) P(x_\alpha; de_\beta).*$$

The following theorem will be useful:

THEOREM 3.4. *Let $\phi(x_{\alpha_1}, \cdots, x_{\alpha_p})$ be a P-measurable integrable function. Then*

$$(3.11)\qquad \int \phi dP = \int P(de_{\alpha_1}) \int P(x_{\alpha_1}; de_{\alpha_1}) \int \cdots \int \phi P(x_{\alpha_1}, \cdots, x_{\alpha_{p-1}}; de_{\alpha_p}).\dagger$$

This theorem can be considered as a corollary to the preceding one, but a direct proof will be given by induction. If $p=1$, (3.11) becomes

$$\int \phi(x_{\alpha_1}) dP = \int \phi(x_{\alpha_1}) P(de_{\alpha_1}),$$

which is certainly true. Suppose that $q>1$ and that the theorem is true for $p<q$. In (3.11) (with $p=q$), the first (symbolic integration) gives $E(x_{\alpha_1}, \cdots, x_{\alpha_{q-1}}; \phi)$. Since the theorem is supposed true for $p=q-1$, the right side of (3.11) then collapses to

$$\int E(x_{\alpha_1}, \cdots, x_{\alpha_{q-1}}; \phi) dP,$$

and this is equal to the left side of (3.11) by the definition of conditional expectation.

Most stochastic processes which have been discussed in the mathematical literature are Markoff processes, that is, processes which satisfy the following

---

* The last expression is only symbolic for the second one unless the conditional expectation concerned can be expressed as an integral. Theorem 3.3 is true and will be used below in a somewhat more general form obtained by considering three groups of subscripts, an α-group, a β-group, and a γ-group, to replace α, β, γ.

† This expression is to be evaluated from right to left. If the expression can be considered as an iterated integral, the theorem becomes the generalization of Fubini's theorem (on the evaluation of a multiple integral by means of iterated integrals) to the most general measures on product spaces on which the field of measurable sets is defined by starting with the sets which are direct products of measurable sets in the component spaces. In the case considered in Fubini's theorem, $P(x_{\alpha_1}, \cdots, x_{\alpha_j}; de_{\alpha_{j+1}}) \equiv P(de_{\alpha_{j+1}}), j \geqq 1$. Lévy (I, p. 73) obtains the generalization (where the product space is n-dimensional euclidean space) by a change of variable which reduces the result to Fubini's theorem.

condition: If $\alpha<\beta$, and if Λ is a P-measurable cylinder set over $x_{\beta+1}, x_{\beta+2}, \cdots$, then, except perhaps for an $(x_\alpha, x_{\alpha+1}, \cdots, x_\beta)$-set of P-measure 0,

$$P(x_\alpha, \cdots, x_\beta; \Lambda) = P(x_\beta; \Lambda).^*$$

It follows at once that if $\alpha_1<\alpha_2< \cdots <\alpha_p$, and if Λ is a P-measurable cylinder set over $x_{\alpha_p+1}, x_{\alpha_p+2}, \cdots$, then if we neglect an $(x_{\alpha_1}, \cdots, x_{\alpha_p})$-set of P-measure 0, $P(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda) \equiv P(x_{\alpha_p}; \Lambda)$. If $\alpha<\beta$, and if Λ is a cylinder set over $x_\beta, x_{\beta+1}, \cdots$, (3.10) implies that

$$P(x_\alpha; \Lambda) = \int P(x_\beta; \Lambda) P(x_\alpha; de_\beta) \tag{3.12}$$

for a Markoff process. Markoff processes are sometimes carelessly discussed in the literature as if they were the general case, as if (3.12) followed from the definition of probability.

4. **Probability measures in terms of the conditional probability functions.** In §3, the conditional probability functions were derived from the measure relations of a stochastic process. In this section the converse problem will be discussed. It will be seen that more is supposed below to be true of the conditional probabilities than is true in the general case, but the hypotheses are wide enough to cover the applications to be made.

Let F_x, F_ω be fields as described in §2. Suppose that for every pair of integers m, n, with $m \leqq n$ and cylinder set Λ_{n+1} in the field F_ω over x_{n+1}, a function $P(x_m, \cdots, x_n; \Lambda_{n+1})$ is defined and has the following properties:

(i) For fixed $x_m, \cdots, x_n$, $P(x_m, \cdots, x_n; \Lambda_{n+1})$ is a probability measure on the field of sets Λ_{n+1}.

(ii) For fixed Λ_{n+1}, $P(x_m, \cdots, x_n; \Lambda_{n+1})$ is measurable with respect to F_ω.

Let $Q(\Lambda)$ be a probability measure defined on the field of cylinder sets of F_ω over x_m. There is then, as we shall now show, a uniquely determined probability measure defined on the cylinder sets of F_ω over $x_m, x_{m+1}, \cdots$, having the given functions $P(x_m, \cdots, x_n; \Lambda_{n+1})$ as its conditional probability functions and equal to $Q(\Lambda)$ if Λ is a cylinder set of F_ω over x_m. If ϕ is the characteristic function of any cylinder set of F_ω over $x_m, \cdots, x_n$, $P(\Lambda)$ is defined by an iterated integral

$$P(\Lambda) = \int Q(de_m) \int P(x_m; de_{m+1}) \int P(x_m, x_{m+1}; de_{m+2}) \int \cdots \int \phi_\Lambda P(x_m, \cdots, x_{n-1}; de_n).^\dagger \tag{4.1}$$

* A detailed discussion of the physical meaning of a Markoff process is given in Kolmogoroff (I).

† Cf. P. Lévy (I, pp. 121–123).

To show that this defines $P(\Lambda)$ uniquely it is necessary to show that if Λ is also a cylinder set over $x_m, \cdots, x_{n'}$, $(n' \neq n)$, the expression (4.1) and the corresponding expression

$$P'(\Lambda) = \int Q(de_m) \int P(x_m;\, de_{m+1}) \int P(x_m,\, x_{m+1};\, de_{m+2}) \int \cdots \int \phi_\Lambda P(x_m, \cdots, x_{n'-1};\, de_{n'}) \tag{4.1'}$$

are equal. We can suppose without restricting generality that $n' > n$. Then

$$P'(\Lambda) = \int Q(de_m) \int \cdots \int P(x_m, \cdots, x_{n-1};\, de_n) \int \cdots \int \phi_\Lambda P(x_m, \cdots, x_{n'-1};\, de_{n'}),$$

and since ϕ_Λ can depend only on $x_m, \cdots, x_n$, the first integration gives $P(x_m, \cdots, x_{n'-1};\, \Omega)$ (which is identically 1) multiplied by ϕ_Λ. Similarly the next integrations, up to the integration over x_n, give ϕ_Λ; hence $P'(\Lambda) = P(\Lambda)$. Evidently $P(\Lambda)$, as thus defined, is a probability measure on the field of cylinder sets of F_ω over any finite set of coordinates with subscripts at least equal to m. It then follows from Theorem 1.1, as applied to the case where $\mathcal{A}$ is the set of integers $m, m+1, \cdots$, and where X, F_x are as here given, that the domain of definition of $P(\Lambda)$ can be extended to include all the cylinder sets of F_ω over $x_m, x_{m+1}, \cdots$ in such a way that the extended set function is a probability measure. Since if Λ, M are respectively cylinder sets over x_{n+1} and $x_m, \cdots, x_n$, with characteristic functions ϕ_Λ, ϕ_M,

$$\int_M P(x_m, \cdots, x_n;\, \Lambda) P(de_{m,\ldots,n}) = \int P(x_m, \cdots, x_n;\, \Lambda) \phi_M P(de_{m,\ldots,n})$$

$$= \int Q(de_m) P(x_m;\, de_{m+1}) \int \cdots \int \phi_M P(x_m, \cdots, x_n;\, \Lambda) P(x_m, \cdots, x_{n-1};\, de_n)$$

$$= \int Q(de_m) \int \cdots \int P(x_m, \cdots, x_{n-1};\, de_n) \int \phi_\Lambda \phi_M P(x_m, \cdots, x_n;\, de_{n+1})$$

$$= P(\Lambda \cdot M);$$

the conditional probability functions determined by $P(\Lambda)$ are actually the given ones. The set function $P(\Lambda)$ thus exists and it is uniquely determined by $Q(E)$ and the given conditional probability functions, since (4.1) holds for $P(\Lambda)$ either as a definition or as a theorem, because of Theorem 3.4.

What made this problem simple was the fact that the given conditional probability functions were entirely independent of each other; that is, there were no necessary relations between the given functions. This was possible because only cylinder sets over $x_m, x_{m+1}, \cdots$ were being considered, *for m fixed*. If m is not to be kept fixed, the set of conditional probability functions can no longer be chosen independently of each other. We shall only consider the problem in detail for conditional probability functions corresponding to Markoff processes. In the treatment just given, if we had supposed that $P(x_m, \cdots, x_n; \Lambda_{n+1})$ depended only on x_n, the resulting process would have been a Markoff process. To extend the results to the consideration of all the sets of F_ω, we shall need the following lemma:

LEMMA 4.1. *Let $\{Q_N(\mathcal{E})\}$ be a sequence of probability measures defined on the sets of some Borel field S of sets of an abstract space. Suppose that*

$$\mathcal{E}_1 \supseteq \mathcal{E}_2 \supseteq \cdots, \quad \prod_1^\infty \mathcal{E}_\nu = 0 \tag{4.2}$$

implies that

$$\lim_{\nu \to \infty} Q_N(\mathcal{E}_\nu) = 0 \tag{4.3}$$

uniformly in N.

(i) *If*

$$\lim_{N \to \infty} Q_N(\mathcal{E}) = Q(\mathcal{E}) \tag{4.4}$$

exists for every $\mathcal{E}$ in S, the set function $Q(\mathcal{E})$ is a probability measure.

(ii) *If S is the Borel field determined by a denumerable collection of its sets, there is a subsequence $\{Q_{N_m}(\mathcal{E})\}$ of $\{Q_N(\mathcal{E})\}$ converging to a limiting probability measure.*

This lemma can be considered as a generalization of Helly's theorem.* Its proof will only be sketched.

Proof of (i). We need only prove that

$$\mathcal{E} = \sum_1^\infty \mathcal{E}_m, \qquad \mathcal{E}_m \cdot \mathcal{E}_n = 0, \qquad (m \neq n), \tag{4.5}$$

implies

$$Q(\mathcal{E}) = \sum_1^\infty Q(\mathcal{E}_m), \tag{4.6}$$

* Sitzungsberichte der Akademie der Wissenschaften, Vienna, class IIa, vol. 121 (1912), p. 286.

or (since $Q(\mathcal{E})$ is obviously additive) that

$$(4.7)\quad Q(\mathcal{E}) - \sum_1^{\nu} Q(\mathcal{E}_m) = Q\Big(\sum_{\nu+1}^{\infty} \mathcal{E}_m\Big) = \lim_{N\to\infty} Q_N\Big(\sum_{\nu+1}^{\infty} \mathcal{E}_m\Big) \to 0, \quad \nu \to \infty.$$

Now

$$\sum_1^{\infty} \mathcal{E}_m \supseteq \sum_2^{\infty} \mathcal{E}_m \supset \cdots, \quad \prod_{N=1}^{\infty} \sum_N^{\infty} \mathcal{E}_m = 0,$$

so that the hypotheses of the lemma imply (4.7).

Proof of (ii). Let $\mathcal{E}_1', \mathcal{E}_2', \cdots$ be a denumerable collection of sets of S, such that the Borel system of sets determined by the sequence $\{\mathcal{E}_m'\}$ is S. By a familiar procedure, we can find a sequence of integers $\{N_n\}$ such that $\lim_{n\to\infty} Q_{N_n}(\mathcal{E})$ exists for every set $\mathcal{E}$ of S which is in the field of sets determined by the sequence of sets $\{\mathcal{E}_n'\}$. The hypotheses then imply that $\lim_{n\to\infty} Q_{N_n}(\mathcal{E})$ exists for every set $\mathcal{E} \in S$,* and the remainder of part (ii) then follows from (i).

THEOREM 4.1. *Suppose that F_x is the Borel field of sets determined by a denumerable subcollection of its sets, and that, for every pair of integers m, n, with $m \leqq n$ and cylinder set Λ in F_ω over x_{n+1}, a (conditional probability) function $P(x_m, \cdots, x_n; \Lambda)$ is defined which has properties* (i), (ii) *given above, and for which also*

$$(4.8)\quad P(x_m, \cdots, x_n; \Lambda) \equiv P(x_n; \Lambda).$$

Suppose that for each fixed value of n, whenever $\Lambda_1, \Lambda_2, \cdots$ is a sequence of cylinder sets of F_ω over x_{n+1} satisfying

$$(4.9)\quad \Lambda_1 \supseteq \Lambda_2 \supseteq \cdots, \quad \prod_1^{\infty} \Lambda_\nu = 0,$$

it is true that

$$(4.10)\quad \lim_{\nu\to\infty} P(x_n; \Lambda_\nu) = 0$$

uniformly in x_n. Then

(i) *there is a Markoff process with the given conditional probability functions, and*

(ii) *if $x_n = x_{n+1}$ implies*

$$(4.11)\quad P(x_n; \Lambda) = P(x_{n+1}; T\Lambda),\dagger$$

---

* This can be proved by transfinite induction.

† The transformation T was defined at the end of §2.

there is a Markoff temporally homogeneous process with the given conditional probability functions.

Proof of (i). Let $Q(E)$ be any probability measure defined on the sets of F_x. Then if M is any cylinder set of F_ω over $x_N, x_{N+1}, \cdots, x_n$ with characteristic function ϕ_M, define $P_N(M)$ by

$$(4.12)\quad P_N(M) = \int Q(de_N) \int P(x_N; de_{N+1}) \int \cdots \int \phi_M P(x_{n-1}; de_n).$$

It was shown above that this determines $P_N(M)$ uniquely. In the following, if M is any cylinder set over a finite number of coordinates $x_m, x_{m+1}, \cdots, x_n$, we define $P(x_{m-1}; M)$ by

$$(4.13)\quad P(x_{m-1}; M) = \int P(x_{m-1}; de_m) \int P(x_m; de_{m+1}) \int \cdots \int \phi_M P(x_{n-1}; de_n),$$

where ϕ_M is the characteristic function of M.

Now let m, n be any two integers with $m \leqq n$. The cylinder sets of F_ω over $x_m, \cdots, x_n$ constitute a Borel field $F_{m,n}$ determined by a denumerable subcollection.* Suppose that $\Lambda_1, \Lambda_2, \cdots$ are sets in the field $F_{m,n}$, and that

$$\Lambda_1 \supseteq \Lambda_2 \supseteq \cdots, \quad \prod_1^\infty \Lambda_\nu = 0.$$

Then

$$(4.14)\quad \lim_{\nu \to \infty} P(x_{m-1}; \Lambda_\nu) = 0,$$

for all x_{m-1}. If $\epsilon > 0$, and if M_ν is the cylinder set over x_{m-1} on which there is a value of $\mu \geqq \nu$ such that $P(x_{m-1}; \Lambda_\mu) > \epsilon$, then it follows from (4.14) that

$$M_1 \supseteq M_2 \supseteq \cdots, \quad \prod_1^\infty M_\nu = 0.$$

It follows from the definition of M_ν and from the fact that the conditional probability functions are less than or equal to 1, that

$$(4.15)\quad P(x_{m-2}; \Lambda_\nu) = \int P(x_{m-1}; \Lambda_\nu) P(x_{m-2}; de_{m-1}) \leqq \epsilon + P(x_{m-2}; M_\nu).$$

The hypotheses of the theorem imply that

$$\lim_{\nu \to \infty} P(x_{m-2}; M_\nu) = 0$$

uniformly in x_{m-2}. If ν_0 is chosen so large that

* Cf. §2.

$$P(x_{m-2};\ \mathrm{M}_\nu) < \epsilon, \quad \nu > \nu_0$$

for all x_{m-2}, it follows from (4.15) that

$$P(x_{m-2};\ \Lambda_\nu) < 2\epsilon, \quad \nu > \nu_0$$

for all x_{m-2}. Then if $N < m-2$, and if $\nu > \nu_0$,

$$P_N(\Lambda_\nu) = \int Q(de_N) \int P(x_N;\ de_{N+1}) \int \cdots \int P(x_{m-2};\ \Lambda_\nu) P(x_{m-3};\ de_{m-2}) < 2\epsilon;$$

so that if the field $F_{m,n}$ is identified with the field S of Lemma 4.1, the hypotheses of the lemma are satisfied. There is therefore a subsequence $\{P_{N_m}(\Lambda)\}$ of $\{P_N(\Lambda)\}$ converging to a limiting probability measure (defined on the field $F_{m,n}$). Since m, n are arbitrary except that $m \leqq n$, there is a further subsequence $\{P_{N_{a_m}}(\Lambda)\}$ such that

$$\lim_{m\to\infty} P_{N_{a_m}}(\Lambda) = P(\Lambda)$$

exists for every cylinder set Λ of F_ω over a finite number of coordinates. Since $P(\Lambda)$ satisfies the hypotheses of Theorem 1.1, its domain of definition can be extended to include all the sets of F_ω. If m, n are again any two integers with $m \leqq n$, and if Λ is a cylinder set of F_ω over x_{n+1}, it was shown in the general discussion preceding the statement of Theorem 4.1, that (if $N \leqq m$),

$$\int_{\mathrm{M}} P(x_n;\ \Lambda) dP_N = \int_{\mathrm{M}} P(x_N, \cdots, x_n;\ \Lambda) dP_N = P_N(\Lambda\, \mathrm{M}),$$

for every set M of $F_{m,n}$; which expresses the fact that the conditional probabilities at the Nth stage are the given ones. If N becomes negatively infinite only assuming values of the sequence $\{N_{a_m}\}$, this becomes

$$\int_{\mathrm{M}} P(x_n;\ \Lambda) dP = P(\Lambda\, \mathrm{M}),$$

so that the conditional probability functions of the new P-measure are the given ones.

Proof of (ii). Suppose that (4.11) is satisfied, and let $Q(E)$ be as in the proof of (i). Let Λ be a cylinder set of F_ω over a finite number of the coordinates $x_2, x_3, \cdots$. Consider the sequence of set functions $\{Q_N(\Lambda)\}$ where

$$Q_N(\Lambda) = \frac{1}{N} \sum_1^N \int P(x_1;\ T^m\Lambda) Q(de_1).$$

A slight modification of the argument just used shows that some subsequence

$\{Q_{N_m}(\Lambda)\}$ of $\{Q_N(\Lambda)\}$ converges for every such set Λ. Moreover, if $P(\Lambda)$ is the limit, then

$$P(T\Lambda) = \lim_{m\to\infty} \frac{1}{N_m} \sum_{1}^{N_m} \int P(x_1; T^{i+1}\Lambda) Q(de_1) = \lim_{m\to\infty} \frac{1}{N_m} \sum_{2}^{N_m+1} \int P(x_1; T^{i}\Lambda) Q(de_1)$$
$$= \lim_{m\to\infty} \frac{1}{N_m} \sum_{1}^{N_m} \int P(x_1; T^{i}\Lambda) Q(de_1) = P(\Lambda),$$

so that if Λ is any cylinder set of F_ω over a finite number of coordinates, we can consistently define $P(\Lambda)$ as $P(T^m\Lambda)$, where m is chosen so large that $T^m\Lambda$ is a cylinder set over $x_2, x_3, \cdots$. The set function thus defined satisfies the conditions of Theorem 1.1, hence it can be extended to become a probability measure defined on all the sets of F_ω. Evidently $P(\Lambda) = P(T\Lambda)$ for every set of F_ω, and, as in the proof of part (i), the conditional probability functions are the given ones. We have proved incidentally the following corollary:

COROLLARY. *In part* (ii) *of Theorem* 4.1, *if $Q(E)$ is any probability measure defined on the field F_x, there is an increasing sequence of positive integers $N_1, N_2, \cdots$ such that*

$$\lim_{\nu\to\infty} \frac{1}{N_\nu} \sum_{1}^{N_\nu} \int P(x_1; T^m\Lambda) Q(de_1) = P(\Lambda)$$

exists for all cylinder sets Λ over a finite number of the coordinates $x_2, x_3, \cdots$ and determines a possible choice of the probability measure $P(\Lambda)$.

The proof becomes particularly simple if a value $x_1^{(0)}$ of x_1 is chosen, and if $Q(E)$ is defined as 1 or 0 according as E does or does not contain $x_1^{(0)}$. The integral then becomes $P(x_1^{(0)}; T^m\Lambda)$.

5. **Examples.** The examples discussed in this section are simple illustrative examples, all of Markoff processes, which will be studied in detail in §7.

I. The type of stochastic process most frequently studied is that in which the chance variables form an independent set, that is, in which if $E_1, \cdots, E_p$ are sets of F_x and $\alpha_1, \cdots, \alpha_p$ are distinct integers, the P-measure of the Ω-set determined by the m conditions $x_{\alpha_j} \in E_j$, $(j = 1, \cdots, m)$, is the product of the P-measures of the m sets determined by the single conditions. This case is characterized by the fact that $P(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda)$ does not depend on $x_{\alpha_1}, \cdots, x_{\alpha_p}$ if Λ is a P-measurable cylinder set over coordinates not including $x_{\alpha_1}, \cdots, x_{\alpha_p}$.*

II. Let X contain n elements, the numbers $1, \cdots, n$. We shall define a

* The corresponding P-measures on Ω have been examined by many writers, referred to in §1 and §2.

Markoff temporally homogeneous process. Let (p_{jk}) be an n^2 matrix of elements which satisfy the conditions

$$(5.1)\qquad \begin{aligned} p_{jk} &\geqq 0, && j, k = 1, \cdots, n, \\ \sum_{k=1}^{n} p_{jk} &= 1, && j = 1, \cdots, n. \end{aligned}$$

The element p_{jk} is identified with the conditional probability that $x_{\nu+1}=k$ if $x_\nu=j$, $\nu=0, \pm 1, \cdots$. The P-measure will be completely determined if the probability p_k that $x_\nu=k$ (which is to be independent of ν) is assigned. The hypotheses of Theorem 4.1 (ii) are satisfied, so the existence of the "absolute probabilities" $p_1, \cdots, p_n$ is assured. These satisfy (cf. equation (3.1))

$$(5.2)\qquad \sum_{j=1}^{n} p_j p_{jk} = p_k, \qquad \sum_{1}^{n} p_j = 1, \qquad k = 1, \cdots, n.$$

According to the corollary to Theorem 4.1, the absolute probabilities can be obtained in the form

$$(5.3)\qquad p_k = \lim_{\nu \to \infty} \frac{1}{N_\nu} \sum_{m=1}^{N_\nu} p_{jk}^{(m)}, \qquad (j \text{ fixed}),$$

where the set $N_1, N_2, \cdots$ is an increasing set of positive integers, and $p_{jk}^{(m)}$ is the conditional probability that $x_{\nu+m}=k$ if $x_\nu=j$. The element $p_{jk}^{(m)}$ is determined (cf. equation (3.12)) by

$$(5.4)\qquad p_{jk}^{(1)} = p_{jk}, \qquad p_{jk}^{(m+1)} = \sum_{l=1}^{n} p_{jl}^{(m)} p_{lk}^{(1)}.$$

Evidently the matrix $(p_{jk}^{(m)})$ is the mth power of the matrix (p_{jk}), and its elements satisfy (5.1).*

III. Let X be arbitrary, but suppose that F_x is the Borel field of sets determined by some denumerable subcollection. A non-negative completely additive set function (not necessarily always finite-valued) is supposed defined on X,† and the integral, with respect to this measure, of an X-measurable

* This classical Markoff process, the one originally studied by Markoff, is discussed by Hostinsky (II), who gives an extensive bibliography. References to more recent work will be given in §7. Fréchet has announced a new book on Markoff processes in which he will probably study this case in detail.

† It is supposed that there is a monotone increasing sequence of sets, each of finite X-measure, whose sum is X, and that the X-measure of any X-measurable set is the limit of the X-measure of its intersection with the sets of the sequence. We shall suppose that this set function is extended as usual so that it is defined (and 0) on the subsets of sets of F_x on which it vanishes. The sets for which the extended set function is defined will be called X-measurable, and X-measurable functions are then defined in the usual way.

function $f(x)$ over an X-measurable set E will be denoted by $\int_E f(x)dx$.*

Let $X \times Y$ be the product space of pairs (x, y), $(x, y \in X)$. A measure can then be defined on $X \times Y$ by the condition that if E, F are X-sets in F_x, the $X \times Y$-measure of the set determined by $x \in E$, $y \in F$ is the product of the X-measures of E and F.† Let $p(x, y)$ be a function defined on $X \times Y$-space which is measurable with respect to the measure just defined,‡ and which satisfies the following conditions:

(a) $p(x, y)$ is non-negative;

(b) $p(x, y)$ is integrable in y for fixed x, and

$$\int p(x, y)dy = 1. \S \tag{5.5}$$

If Λ is a cylinder set over $x_{\nu+1}$ determined by the condition $x_{\nu+1} \in E$, $(E \in F_x)$, we define $P(x_\nu; \Lambda)$ by

$$P(x_\nu; \Lambda) = \int_E p(x, y)dy, \qquad \nu = 0, \pm 1, \cdots. \tag{5.6}$$

By Theorem 4.1 (ii) these conditional probability functions are those of a temporally homogeneous Markoff process if, whenever $E_1, E_2, \cdots$ are sets in the field,

$$E_1 \supseteq E_2 \supseteq \cdots, \quad \prod_1^\infty E_m = 0 \tag{5.7}$$

implies

$$\lim_{m \to \infty} \int_{E_m} p(x, y)dy = 0 \tag{5.8}$$

uniformly in x. This can be interpreted as the uniform (in x) integrability of $p(x, y)$ with respect to y, that is, the uniform (in x) absolute continuity (in E)

* It will be supposed, as usual, that integrability means absolute integrability, and that a non-negative function is integrable if and only if its integral on the sequence of sets in the preceding note is bounded. In many applications, X is supposed to be a Borel set $\mathcal{E}$ of n-dimensional euclidean space and F_x the field of Borel subsets of $\mathcal{E}$; and the set function is supposed to be Borel measure.

† Saks, *Théorie de l'Intégrale*, Warsaw, 1933, pp. 257–263. As usual we suppose that the measure is further extended so that subsets of sets of measure 0 are measurable and of measure 0.

‡ We shall suppose further that $p(x, y)$ is measurable with respect to the $X \times Y$ measure, as defined before its extension described in the preceding note, so that $p(x_0, x_1)$ is measurable with respect to F_ω, considered as a function defined on Ω. In any case $p(x, y)$ will be equal to such a function almost everywhere on $X \times Y$-space. It then follows (cf. Saks, ibid., p. 258) that $p(x, y)$ is X-measurable in x (y) for each fixed value of y (x).

§ As in the previous sections, when no region of integration is explicitly prescribed integration will be over the whole space.

of the set function $\int_E p(x, y)dy$. The condition will be satisfied if $p(x, y) \leqq \phi(y)$, for all x, y, where $\phi(x)$ is X-measurable and integrable over X. If the X-measure of X is finite, that is, if $\int 1dx < \infty$, the condition will be satisfied if $p(x, y)^2$ is integrable in y, and if there is a number K such that for every value of x,

$$\int p(x, y)^2 dy \leqq K.$$

If (5.7) implies (5.8), the measure function $P(\Lambda)$ given by Theorem 4.1 becomes, on the cylinder sets over x_1, a function $Q(E)$ of sets $E \in F_x$; and if Λ is determined by the condition $x_1 \in E$, $Q(E) = P(\Lambda)$. Moreover (cf. equation (3.1)),

$$\int Q(dx) \int_E p(x, y)dy = Q(E). \tag{5.9}$$

If the X-measure of E vanishes, (5.9) shows that $Q(E) = 0$, that is, $Q(E)$ is absolutely continuous. There is then an X-measurable function $p(x)$ for which

$$Q(E) = \int_E p(x)dx,$$

for all sets $E \in F_x$.* This function $p(x)$ satisfies the equation

$$\int p(x)dx \int_E p(x, y)dy = \int_E p(y)dy, \tag{5.9'}$$

so that, if the order of integration is interchanged ($p(x, y)$, $p(x)$ are non-negative),

$$\int_E dy \int p(x)p(x, y)dx = \int_E p(y)dy.$$

Then

$$\int \left\{ \int p(x)[p(x, y) - p(y)]dx \right\} dy = 0. \tag{5.10}$$

Since E is arbitrary, (5.10) implies

$$\int p(x)p(x, y)dx = p(y) \tag{5.11}$$

for almost all y. The function $p(y)$ can now be changed on a set of X-measure 0 to make (5.11) true for all y. According to the corollary to Theorem 4.1, the "absolute probability density" $p(y)$ can be obtained in the form

* Saks, *Théorie de l'Intégrale*, Warsaw, 1933, p. 257.

$$\int_E p(y)dy = \lim_{\nu\to\infty} \frac{1}{N_\nu} \sum_1^{N_\nu} \int_E p^{(m)}(x, y)dy, \qquad (x \text{ fixed}), \tag{5.12}$$

where $N_1, N_2, \cdots$ is an increasing sequence of positive integers, and $\int_E p^{(m)}(x, y)dy$ is the conditional probability that $x_{\nu+m} \in E$ if $x_\nu = x$. The function $p^{(m)}(x, y)$ is determined (cf. equation (3.12)) by

$$p^{(1)}(x, y) = p(x, y), \qquad p^{(m+1)}(x, y) = \int p^{(m)}(x, z)p(z, y)dz. \tag{5.13}$$

Evidently the function $p^{(m)}(x, y)$ satisfies the conditions (a), (b) imposed on $p(x, y)$. Moreover if the sequence of sets satisfies (5.7), and if (5.7) implies (5.8), then

$$\begin{aligned}\int_{E_\nu} p^{(m)}(x, y)dy &= \int_{E_\nu} dy \int p^{(m-1)}(x, z)p(z, y)dz \\ &= \int p^{(m-1)}(x, z)dz \int_{E_\nu} p(z, y)dy \to 0,\end{aligned}$$

uniformly in x, so that $p^{(m)}(x, y)$ satisfies the condition of uniform integrability if $p(x, y)$ does. A slight modification of the proof shows that if, for some integer $\mu \geqq 1$, $p^{(\mu)}(x, y)$ satisfies the condition of uniform integrability, the function $p^{(m)}(x, y)$ for $m > \mu$ will also satisfy the condition; and then a suitable modification of the proof of Theorem 4.1 (ii) will show that *it is sufficient to assure the existence of an absolute probability density (given, for example, by* (5.12)) *to suppose that for some* m, $p^{(m)}(x, y)$ *satisfies the uniform integrability condition.*

The Markoff process considered here is very general.* Example II is a special case. To show this we need only define X-measure suitably and define the function $p(x, y)$ in terms of p_{jk}. The space X has points denoted by the numbers $1, \cdots, n$. If E is any set of X containing r elements, define the X-measure of E as r. The function $p(x, y)$ is defined as p_{jk} for $x = j$, $y = k$. More generally we can consider a space X whose points are the numbers $1, 2, \cdots$. The field F_x is to be the field of all subsets of X, and the X-measure of a set is the number of points in it. A matrix (p_{jk}) is given whose elements satisfy

$$p_{jk} \geqq 0, \qquad \sum_{k=1}^{\infty} p_{jk} = 1. \tag{5.14}$$

* Hostinsky discusses this type in (I) and (II) and gives an extensive bibliography in (II). Cf. also the forthcoming book by Fréchet. Further references will be given in §7.

The condition of uniform integrability becomes here the condition of uniform convergence in (5.14):

$$\lim_{N\to\infty} \sum_{k=1}^{N} p_{jk} = 1 \tag{5.15}$$

uniformly in j. If the condition is satisfied, absolute probabilities $p_1, p_2, \cdots$ exist satisfying the conditions

$$p_j \geqq 0, \quad j \geqq 1; \qquad \sum_{j=1}^{\infty} p_j p_{jk} = p_k, \quad k \geqq 1; \qquad \sum_{1}^{\infty} p_j = 1. \tag{5.16}$$

Let $q_0, q_1, \cdots$ be a sequence of non-negative numbers whose sum is 1. Suppose that $p_{jk}=0$ if $k<j$, and $p_{jk}=q_{k-j}$ if $k\geqq j$. If $q_0<1$, it is readily seen that no process exists, temporally homogeneous or not, having the given conditional probabilities.* A particular case in which this is obvious is obtained by setting $q_1=1$.

IV. The following example is again that of a temporally homogeneous Markoff process. The space X is arbitrary, but we suppose that a probability measure is defined on the field F_x, and that a transformation Sx is defined on X which is one-to-one, takes X-measurable sets into X-measurable sets, and is X-measure preserving. If Λ is a cylinder set of F_ω over $x_{\nu+1}$, determined by the condition $x_{\nu+1} \in E$, define $P(x_\nu; \Lambda)$ by

$$P(x_\nu; \Lambda) = 1 \quad \text{if} \quad Sx_\nu \in E,$$
$$P(x_\nu; \Lambda) = 0 \quad \text{if} \quad Sx_\nu \notin E.$$

The condition of Theorem 4.1 is not satisfied, but there is nevertheless a temporally homogeneous Markoff process with these conditional probability functions, for if M is an Ω-set determined by the conditions $x_{\alpha_j} \in E_j$, $(j=1, \cdots, p)$, the P-measure of M can be defined as the X-measure of the set $(S^{-\alpha_1}E_1)(S^{-\alpha_2}E_2) \cdots (S^{-\alpha_p}E_p)$.

6. **Temporally homogeneous processes.** A stochastic process suggests the transformation idea in its very phraseology; for example, "the conditional probability that x_1 belong to a set E if $x_0=x_0^{(0)}$"; and in §1 an explicit transformation T was defined to exploit this suggestion. In this section we shall consider only temporally homogeneous processes (for which T is measure-preserving). The theory of temporally homogeneous processes uses to a large extent the terminology of the theory of measure-preserving transformations, and in this section we shall see that this has a complete justification, in every

* This statement refers only to a process corresponding to a sequence of chance variables $\cdots, x_{-1}, x_0, x_1, \cdots$. The results of the preceding section show that the statement is not true if processes corresponding to a sequence of chance variables $x_1, x_2, \cdots$ are being considered.

detail, through the mediation of the transformation T.† The present section will then essentially be an independent study of the measure-preserving transformation T, with particular stress on the case where the P-measure satisfies the conditions imposed on the P-measure corresponding to a Markoff process. We shall apply the theory of measure-preserving transformations, as developed by Birkhoff, Koopman, and von Neumann.

Suppose that a given process is temporally homogeneous. The ergodic theorem gives the following result.‡

THEOREM 6.1. *Let the given process be temporally homogeneous, and let* Λ *be any* P*-measurable set.*

(i) *If* $\phi(\omega)$ *is* P*-measurable and integrable, there is a* P*-measurable function* $\phi^*(\omega)$ *such that*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{1}^{N} \phi(T^m\omega) = \phi^*(\omega) \tag{6.1}$$

almost everywhere on Ω. *In particular there is a* P*-measurable function* $Q(\omega; \Lambda)$ *such that*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{1}^{N} P(x_m; T^m\Lambda) = Q(\omega; \Lambda) \tag{6.2}$$

almost everywhere on Ω.

(ii) *If the process is a Markoff process, and if* Λ *is any* P*-measurable cylinder set over* $x_\nu, x_{\nu+1}, \cdots$ *for some integer* ν,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{1}^{N} P(x_0; T^m\Lambda) = P^*(x_0; \Lambda) \tag{6.3}$$

exists almost everywhere on Ω; *that is, except possibly on an* x_0*-set of* P*-measure* 0.

The fact that the limit exists in (6.2) is apparently new. Results closely related to (ii), with more restrictive hypotheses on the conditional probability functions,§ have been proved by Fréchet and (jointly) by Kryloff and Bogolioùboff.

† Conversely, as was seen in example IV, a measure-preserving transformation gives rise to a certain (Markoff) temporally homogeneous process, which is necessarily of a very special type.

‡ The form of the ergodic theorem used here (due to Birkhoff) is the following: If $T\omega$ is a measure-preserving transformation of an abstract space Ω, then part (i) of the following theorem (Theorem 6.1) holds. A simple proof was given by Khintchine, Mathematische Annalen, vol. 107 (1933), pp. 485–488. The function $\phi(x, N)/N$ of Khintchine's proof corresponds to the average in (6.1). For a complete treatment of the ergodic and related theorems see E. Hopf, *Ergodentheorie*, Ergebnisse der Mathematik, vol. 5, no. 2, which appeared so late that detailed reference to it could not be made in this paper.

§ The only restriction on the conditional probability functions made in Theorem 6.1 is that there should actually exist a corresponding temporally homogeneous process; that is, that there should exist "absolute probabilities." Exact references will be given in §7.

Proof of (i). The first part of (i) is a restatement of the ergodic theorem. The second part of (i) is an application of the first part, with $\phi(\omega) = P(x_0; \Lambda)$.

Proof of (ii). Suppose that Λ is as described in (ii). Then if $m > -\nu$, it follows from equation (3.12) that

$$\int P(x_m; T^m\Lambda)P(x_0; de) = \int P(x_m; T^m\Lambda)P(x_0; de_m) = P(x_0; T^m\Lambda),$$

neglecting sets of P-measure 0, so that

$$\int \left\{ \lim_{N\to\infty} \frac{1}{N} \sum_1^N P(x_m; T^m\Lambda) \right\} P(x_0; de) = \lim_{N\to\infty} \frac{1}{N} \sum_1^N \int P(x_m; T^m\Lambda)P(x_0; de)$$

$$= \lim_{N\to\infty} \frac{1}{N} \sum_{|\nu|+1}^N P(x_0; T^m\Lambda) = \lim_{N\to\infty} \frac{1}{N} \sum_1^N P(x_0; T^m\Lambda),$$

neglecting x_0-sets of P-measure 0.†

COROLLARY 1. *If the transformation T is metrically transitive,‡ then*

$$\phi^*(\omega) \equiv \int \phi(\omega) dP, \quad Q(\omega; \Lambda) = P^*(x_0; \Lambda) = P(\Lambda)$$

almost everywhere on Ω.

This corollary is merely a rephrasing, pertinent to the case being considered, of the ergodic theorem for metrically transitive systems.§

COROLLARY 2. *If F_x is the Borel field determined by a denumerable collection of its sets, and if there are no angle variables,‖ then*

---

† If the first two of the above expressions are considered as actual integrals, the admissibility of the transition from the first to the second follows from Lebesgue's theorem on the admissibility of term by term integration of a uniformly bounded convergent sequence of measurable functions. However, even if the integrals are considered merely as symbols for conditional expectations, the proof of Lebesgue's theorem can be extended to this case.

‡ Metric transitivity means here that no P-measurable set of measure not 0 or 1 is invariant under T. If there is a P-measurable set Λ of measure not 0 or 1, which is invariant neglecting a set of P-measure 0, that is, if $T\Lambda = \Lambda + \Lambda_0 - \Lambda_0'$, where $P(\Lambda_0) = P(\Lambda_0') = 0$, then the P-measurable set $\sum_{-\infty}^{\infty} T^m\Lambda$ has measure $P(\Lambda) \neq 0, 1$ and is invariant under T, so there cannot be metric transitivity. Hence the content of the definition is not changed if invariance up to a set of P-measure 0 is substituted for actual invariance.

§ Cf. Khintchine, ibid., p. 488.

‖ An angle variable is a complex-valued P-measurable function $\phi(\omega)$ such that $|\phi| > 0$ on an Ω-set of positive P-measure, and that

$$\phi(T\omega) = c\phi(\omega), \qquad (|c| = 1, c \neq 1),$$

almost everywhere on Ω. If the transformation is metrically transitive, the invariance of $|\phi(\omega)|$ under T implies that $|\phi| \equiv$ const. almost everywhere on Ω. (Cf. B. O. Koopman, Proceedings of the National Academy of Sciences, vol. 17 (1935), pp. 315–318.)

$$\lim_{m\to\infty} P(\mathrm{M}T^m\Lambda) \tag{6.4}$$

exists, when m is restricted to a certain increasing set of integers of measure 1,* *independent of the sets* Λ, M *which can be any P-measurable sets. If there is also metric transitivity, the limit in* (6.4) *is* $P(\Lambda)\cdot P(\mathrm{M})$.

Conversely, if the limit in (6.4) *exists, for all P-measurable sets* Λ, M *on some set of integers of measure* 1, *there are no angle variables; and if the limit is* $P(\Lambda)\cdot P(\mathrm{M})$, *there is metric transitivity.*

This theorem was proved by Koopman and von Neumann in the metrically transitive case, for a one-parameter family of transformations $\{T_t\}$, $-\infty<t<\infty$.† Their proof is applicable, with insignificant modifications, to the family, considered here, of transformations $T_n=T^n$.‡

LEMMA 6.1. *Let* $f(\omega)$ *be any complex-valued P-measurable function, and let* $m_1, m_2, \cdots$ *be an increasing sequence of positive integers. Suppose that* $\{f(T^{m_i}\omega)\}$ *and* $\{f(T^{-m_i}\omega)\}$ *are sequences of functions convergent almost everywhere on* Ω. *Then if O is any open set of the complex plane, the* Ω*-sets defined by the conditions*

$$\lim_{j\to\infty} f(T^{m_j}\omega)\in O, \qquad \lim_{j\to\infty} f(T^{-m_j}\omega)\in O \tag{6.5}$$

are respectively cylinder sets over $x_{-1}, x_{-2}, \cdots$ *and* $x_1, x_2, \cdots$ *(if we neglect sets of P-measure* 0).

To any positive integer ν corresponds (cf. §2) a P-measurable function $f_\nu(\omega)$ depending on only a finite number of coordinates, and having the property that

$$|f(\omega)-f_\nu(\omega)|<1/\nu \tag{6.6}$$

except perhaps on an Ω-set of P-measure at most $2^{-\nu}$. There is a subsequence

* A set of integers $a_1, a_2, \cdots$, $(a_1<a_2<\cdots)$, is said to have measure 1 if

$$\lim_{m\to\infty}\frac{1}{m}\sum_{a_j\leq m} 1=1.$$

† Proceedings of the National Academy of Sciences, vol. 17 (1935), pp. 315–318. To extend their proof to the non-metrically transitive case, it is only necessary to allow a wider interpretation of their projection operator E_0. The hypotheses of topological character they impose on their space are unnecessary in this application.

‡ We use the fact that the P-measurable complex-valued function whose absolute values squared are integrable, form a unitary space H when distance and inner product are defined in the usual way (cf. Stone, *Linear Transformations in Hilbert Space*, American Mathematical Society Colloquium Publications, vol. 15, pp. 23–29). The hypothesis that F_x is the Borel field determined by a denumerable collection of its sets means that H is separable and thus is either finite-dimensional or a Hilbert space. The theorem (and proof) of Koopman and von Neumann is valid in the finite case also.

$\{\mu_j\}$ of $\{m_j\}$ such that for each positive integer j, $f_j(T^{\mu_j}\omega)$ $(f_j(T^{-\mu_j}\omega))$ depends only on $x_{-1}, x_{-2}, \cdots$ $(x_1, x_2, \cdots)$. Since T is measure-preserving,

$$\begin{aligned} &|f(T^{\mu_j}\omega) - f_j(T^{\mu_j}\omega)| < 1/j \\ &(|f(T^{-\mu_j}\omega) - f_j(T^{-\mu_j}\omega)| < 1/j), \end{aligned} \tag{6.7}$$

(j fixed, except perhaps on an Ω-set of P-measure at most 2^{-j}. Then

$$\begin{aligned} &|f(T^{\mu_\nu}\omega) - f_\nu(T^{\mu_\nu}\omega)| < 1/N, \qquad \nu = N, N+1, \cdots \\ &(|f(T^{-\mu_\nu}\omega) - f_\nu(T^{-\mu_\nu}\omega)| < 1/N, \quad \nu = N, N+1, \cdots) \end{aligned} \tag{6.8}$$

except perhaps on a set of P-measure at most 2^{-N+1}; so that

$$\begin{aligned} &\lim_{\nu\to\infty} f_\nu(T^{\mu_\nu}\omega) = \lim_{\nu\to\infty} f(T^{\mu_\nu}\omega) \\ &\left(\lim_{\nu\to\infty} f_\nu(T^{-\mu_\nu}\omega) = \lim_{\nu\to\infty} f(T^{-\mu_\nu}\omega)\right) \end{aligned} \tag{6.9}$$

almost everywhere on Ω. Then the sets defined by (6.5) are the same as the sets defined by the conditions

$$\lim_{\nu\to\infty} f_\nu(T^{\mu_\nu}\omega) \in O \qquad \lim_{\nu\to\infty} f_\nu(T^{-\mu_\nu}\omega) \in O$$

respectively, if we neglect sets of P-measure 0. This fact implies the truth of the lemma.

LEMMA 6.2. *The equality*

$$P(\cdots, x_{n-1}, x_n; \Lambda) = P(x_n; \Lambda)^*$$

holds almost everywhere on the space Ω of a Markoff process, for any P-measurable cylinder set Λ over $x_{n+1}, x_{n+2}, \cdots$.

We need only show that, if M is a P-measurable cylinder set over $\cdots, x_{n-1}, x_n$, then

$$\int_M P(x_n; \Lambda)dP = P(\Lambda M). \tag{6.10}$$

It is evidently sufficient to prove (6.10) for sets M which are cylinder sets over a finite number of coordinates. If M is such a cylinder set, over $x_m, x_{m+1}, \cdots, x_n$, (6.10) follows from the fact that $P(x_m, \cdots, x_n; \Lambda) \equiv P(x_n; \Lambda)$ almost everywhere on Ω.

* The conditional probability function $P(\cdots, x_{n-1}, x_n; \Lambda)$ is defined in the same way as the function $P(x_m, \cdots, x_n; \Lambda)$. Note that the finiteness of the set $\alpha_1, \cdots, \alpha_p$ was not used in the definition of the latter function.

THEOREM 6.2. *A temporally homogeneous Markoff process is metrically transitive* if and only if there is no set E of the field F_x such that if Λ is the Ω-set determined by the condition $x_1 \in E$, $(0<P(\Lambda)<1)$, and*

(i) *Λ is invariant under T, if we neglect a set of P-measure 0, or*

(ii) *if we neglect x_0-sets of P-measure 0,*

then

$$\begin{aligned} P(x_0; \Lambda) &= 1, \qquad x_0 \in E \\ P(x_0; \Lambda) &= 0, \qquad x_0 \notin E. \end{aligned} \tag{6.11}$$

Proof of (i). If there is an invariant set Λ of the type described in the theorem, the process cannot be metrically transitive, by the definition of metric transitivity. Conversely, if the process is not metrically transitive, there is a P-measurable set M invariant under T, and $0<P(\mathrm{M})<1$. If $f(\omega)$ is the characteristic function of M, $f(\omega)=f(T\omega)$ on Ω so that

$$\lim_{m\to\infty} f(T^m\omega) = \lim_{m\to\infty} f(T^{-m}\omega) = f(\omega) \tag{6.12}$$

on Ω. Then by Lemma 6.1, M can be considered either as a cylinder set over $x_{-1}, x_{-2}, \cdots$ (when we denote it by M_1), or as a cylinder set over $x_1, x_2, \cdots$ (when we denote it by M_2), neglecting sets of P-measure 0. It follows from

$$\begin{aligned} &0 \leqq P(\cdots, x_{-1}, x_0; \mathrm{M}_2) \leqq 1, \\ &\int_{\mathrm{M}_1} P(\cdots, x_{-1}, x_0; \mathrm{M}_2)\,dP = P(\mathrm{M}_1\mathrm{M}_2) = P(\mathrm{M}_1), \\ &\int_{C\mathrm{M}_1} P(\cdots, x_{-1}, x_0; \mathrm{M}_2)\,dP = P(C\mathrm{M}_1\mathrm{M}_2) = 0 \end{aligned} \tag{6.13}$$

that

$$\begin{aligned} P(\cdots, x_{-1}, x_0; \mathrm{M}_2) &= 1, \qquad \omega \in \mathrm{M}_1, \\ P(\cdots, x_{-1}, x_0; \mathrm{M}_2) &= 0, \qquad \omega \notin \mathrm{M}_1, \end{aligned} \tag{6.14}$$

if we neglect sets of P-measure 0. Now, according to Lemma 6.2, $P(\cdots, x_{-1}, x_0; \mathrm{M}_2)=P(x_0; \mathrm{M}_2)$ almost everywhere on Ω. Then if a set of P-measure 0 is neglected, M must be a cylinder set over x_0; this set is determined by the condition $x_0 \in E$, where E is the x_0-set on which $P(x_0; \mathrm{M}_2)=1$. Since we can suppose (cf. §2), altering $P(x_0; \mathrm{M}_2)$ on an x_0-set of P-measure 0 if necessary, that $P(x_0; \mathrm{M}_2)$ is measurable with respect to F_ω, we can suppose that E is in F_x. The Ω-set determined by the condition $x_0 \in E$ is invariant (up

* If the transformation T is metrically transitive, or has angle variables, the same will be said to be true of the corresponding stochastic process.

to a set of P-measure 0); so that it is the same as any set determined by a condition $x_n \epsilon E$, up to a set of P-measure 0.

Proof of (ii). If there is a set E, as described in the theorem, for which the hypotheses of (ii) are true, then

$$P(T^{-1}\Lambda; \Lambda) = \int_{T^{-1}\Lambda} P(x_0; \Lambda) dP = P(\Lambda),$$

so that $T^{-1}\Lambda$, and therefore Λ, is invariant under T up to a set of P-measure 0; and the process cannot be metrically transitive. Conversely, if the process is not metrically transitive, an invariant set Λ of the type described in part (i) exists, and (6.14) becomes precisely (6.11).

THEOREM 6.3. *Suppose that $\phi(\omega)$ is an angle variable of a temporally homogeneous Markoff process, so that*

$$\phi(T\omega) = c\phi(\omega), \qquad |c| = 1,\ c \neq 1, \tag{6.15}$$

almost everywhere on Ω.

(i) *The function $\phi(\omega)$ can be considered as a function of x_0 alone, namely, $\phi(\omega) = \psi(x_0)$, so that* (6.15) *becomes*

$$\psi(x_1) = c\psi(x_0), \tag{6.15'}$$

and the possible exceptional set is an (x_0, x_1)-set of P-measure 0.

(ii) *If the hypotheses of Theorem* 3.2 *are satisfied, and if the conditional probability functions are supposed defined as described in the statement of that theorem, then for each value of x_0,*

$$\psi(x_1) = \text{const.} = c\psi(x_0) \tag{6.16}$$

on a cylinder set $\Lambda(x_0)$ over x_1 such that $P(x_0; \Lambda(x_0)) = 1$ except possibly on an x_0-set of P-measure 0.

(iii) *If $\psi(x_0)$ takes on any non-zero value on a set of positive P-measure, c is a root of unity.*

(iv) *There exist P-measurable cylinder sets Λ_0, Λ_1 over x_0, x_1 respectively, determined by the conditions $x_0 \epsilon E_0$, $x_1 \epsilon E_1$, such that $0 < P(\Lambda_i) < 1$, and if we neglect x-sets of P-measure* 0,

$$\begin{aligned} P(x_0; \Lambda_1) &= 1, \qquad x_0 \epsilon E_0, \\ P(x_0; \Lambda_1) &= 0, \qquad x_0 \notin E_0. \end{aligned} \tag{6.17}$$

(v) *The function $\psi(x_0)$, if it is integrable, satisfies the integral equation*

$$\int \psi(x_1) P(x_0; de_1) = c\psi(x_0), \tag{6.18}$$

except possibly for an x_0-set of P-measure 0.*

Proof of (i). Suppose that (6.15) is satisfied. There is an increasing sequence of positive integers $n_1, n_2, \cdots$ such that

$$\lim_{\nu\to\infty} c^{n_\nu} = \lim_{\nu\to\infty} c^{-n_\nu} = 1.$$

Then

$$\lim_{\nu\to\infty} \phi(T^{n_\nu}\omega) = \lim_{\nu\to\infty} \phi(T^{-n_\nu}\omega) = \phi(\omega), \tag{6.19}$$

almost everywhere on Ω. If $\Lambda(O)$ is the Ω-set determined by the condition $\phi(\omega) \in O$ (O an open set of the complex ϕ-plane), the method used in the proof of the preceding theorem shows that $\Lambda(O)$ is a cylinder set over x_0, neglecting an Ω-set of P-measure 0. It follows readily from this that there is a P-measurable function $\psi(x_0)$, depending only on x_0 and such that $\phi(\omega) = \psi(x_0)$ almost everywhere on Ω.

Proof of (ii). Let $\Lambda(x_0)$ be the cylinder set over x_1, determined by the condition $x_1 \in E(x_0)$, on which $\psi(x_1) = c\psi(x_0)$. The (x_0, x_1)-set M, determined by the condition that $x_1 \in E(x_0)$ for each value of x_0, is of P-measure 1, and its measure can be expressed, according to Theorem 3.4, as

$$\int P(de_0) \int f_M P(x_0; de_1) = \int P(x_0; \Lambda(x_0)) P(de_0) = 1,$$

where f_M is the characteristic function of M. Then $P(x_0; \Lambda(x_0)) = 1$ except possibly on an x_0-set of P-measure 0.

Proof of (iii). If $\psi(x_0)$ takes on a value $\psi_0 \neq 0$ on a set Λ_0 of positive P-measure, $\psi(x_0)$ must take on $c^n\psi_0$ on a set Λ_n of the same P-measure (since $\psi(x_n) = c^n\psi(x_0)$). The number c must then be a root of unity; for if not, the numbers $\psi_0, c\psi_0, \cdots$ are all distinct, so that the sets $\Lambda_0, \Lambda_1, \cdots$ are all disjunct. But this is impossible since then

$$1 \geqq P\left(\sum_0^{n-1} \Lambda_m\right) = nP(\Lambda_0), \qquad n = 1, 2, \cdots.$$

Evidently the fact that the process is a Markoff process was not needed in this proof of (iv).

Proof of (iv). If O is an open set of the complex ψ-plane, so chosen that the P-measure of the Ω-set Λ_0, determined by the condition $\psi(x_0) \in O$, is posi-

* The left side of (6.18) is the conditional expectation $E(x_0; \psi)$, and the conditions under which it can be considered an integral were considered in §3. The integrability condition imposed on ψ is unimportant, since if ψ is an angle variable, the integrable function ψ_K, equal to ψ if $|\psi| \leqq K$ and otherwise equal to K, satisfies (6.15′) almost everywhere, so that ψ_K is also an angle variable if K is chosen so large that $|\psi_K| > 0$ on a set of positive P-measure.

tive and less than 1; and if Λ_1 is the Ω-set determined by the condition $c^{-1}\psi(x_1) \epsilon O$, then $\Lambda_0 = \Lambda_1$ (if we neglect sets of P-measure 0). Equation (6.17) then follows at once. (Cf. the proof of Theorem 6.2 (ii).)

Proof of (v). If $\psi(x_1) = c\psi(x_0)$ almost everywhere on Ω, and if ψ is integrable, then

$$\int \psi(x_1)P(x_0; de_1) = c \int \psi(x_0)P(x_0; de_1) = c\psi(x_0),$$

except possibly on an x_0-set of P-measure 0, as was to be proved.

THEOREM 6.4. *A temporally homogeneous process for which the corresponding sequence of chance variables* $\cdots, \mathbf{x}_{-1}, \mathbf{x}_0, \mathbf{x}_1, \cdots$ *form an independent set (cf.* §5, *example* I) *is metrically transitive and has no angle variables.**

A process of this type is a very special case of a Markoff process, so Theorems 6.2 and 6.3 are applicable. A set Λ, as described in the statement of Theorem 6.2, is impossible, since in the case of independence $P(x_0; \Lambda) \equiv P(\Lambda)$; and the process is therefore metrically transitive. For the same reason, there can be no sets Λ_0, Λ_1, as described in Theorem 6.2 (ii); and the process therefore has no angle variables.

THEOREM 6.5. *Suppose the measure relations of a temporally homogeneous Markoff process have the following property: There is a function* $\phi(x_0; x_1)$, *measurable with respect to* F_ω *and integrable over* Ω *in* x_1 *for fixed* x_0, *such that (except possibly for an* x_0*-set of* P*-measure* 0),

$$P(x_0; \Lambda) = \int_\Lambda \phi(x_0; x_1)P(de_1) \tag{6.20}$$

whenever Λ *is a cylinder set of* F_ω *over* x_1.

(i) *The process has no angle variables if and only if* $\lim_{m\to\infty} P(x_0; T^m\Lambda)$ *exists (except possibly on an* x_0*-set of* P*-measure* 0) *for every such set* Λ.

(ii) *The process has no angle variables and is metrically transitive if and only if* $\lim_{m\to\infty} P(x_0; T^m\Lambda) = P(\Lambda)$ *(except possibly on an* x_0*-set of* P*-measure* 0) *for every such set* Λ.

(iii) *If it is true that* $P(x_0; \Lambda)$ *(for* Λ *a cylinder set of* F_ω *over* x_1*) can be defined to be a probability measure, for each fixed value of* x_0, *which vanishes identically in* x_0 *for a given set* Λ *if it vanishes at all, then a function* ϕ *exists and satisfies the hypotheses of the theorem.*

Before proving the theorem, we shall give an example of a temporally homogeneous stochastic process which has no angle variables, and for which

* This result was proved by Doob (I, pp. 761–763), and Hopf (I, p. 95).

the limit described in (i) does not exist. This example will therefore show that some condition, such as the existence of the function ϕ as described, is necessary in the theorem. The example is a particular case of example IV of §5. In example IV suppose that the transformation S is metrically transitive and has no angle variable. Then it is readily seen that the transformation T is metrically transitive and has no angle variable. On the other hand, if Λ is a cylinder set of F_ω over x_1, determined by the condition $x_1 \epsilon E$, then $P(x_0; T^m\Lambda)$ is 1 or 0 according as $S^m x_0$ is or is not in E. Then $\lim_{m\to\infty} P(x_0; T^m\Lambda)$ exists almost everywhere on Ω only if $S^m x_0$, for large m, is finally always in E or never in E for almost all x_0. This implies that E is invariant under S (up to a set of X-measure 0), which is impossible, since S is metrically transitive, if we choose Λ so that $P(\Lambda) \neq 0, 1$.

Proof of (i). Suppose first that there are no angle variables and that F_x is the Borel field determined by a denumerable collection of its sets. According to Theorem 6.1, Corollary 2, there is an increasing sequence of integers, independent of Λ, M, such that

$$\lim_{\nu\to\infty} P(\mathrm{M}T^{a_\nu}\Lambda) = \lim_{\nu\to\infty} \int_{\mathrm{M}} P(x_1; T^{a_\nu}\Lambda)dP = Q(\mathrm{M}; \Lambda) \quad . \tag{6.21}$$

exists, where Λ, M are P-measurable cylinder sets over x_1. From this it follows readily that if $f(x_1)$ is P-measurable and integrable over Ω, then

$$\lim_{\nu\to\infty} \int P(x_1; T^{a_\nu}\Lambda)f(x_1)dP = \int f(x_1)Q(de_1; \Lambda).^* \tag{6.22}$$

In particular (cf. equation (3.12))

$$\begin{aligned} \lim_{\nu\to\infty} \int P(x_1; T^{a_\nu}\Lambda)\phi(x_0, x_1)P(de_1) &= \lim_{\nu\to\infty} \int P(x_1; T^{a_\nu}\Lambda)P(x_0; de_1) \\ &= \lim_{\nu\to\infty} P(x_0; T^{a_\nu}\Lambda) \end{aligned} \tag{6.23}$$

exists. We shall denote this limit by $Q(x_0; \Lambda)$. Evidently

$$\int_{\mathrm{K}} Q(x_0; \Lambda)dP = Q(\mathrm{K}; \Lambda) \tag{6.24}$$

if K is a P-measurable cylinder set over the coordinates $x_m, x_{m+1}, \cdots, x_0$ $(m \leqq 0)$. If $\epsilon > 0$, there is an integer $N = N(\epsilon)$ so large that if $a_j > N$,

$$| P(x_0; T^{a_j}\Lambda) - Q(x_0; \Lambda) | \leqq \epsilon/6$$

* The set function $Q(\mathrm{M}; \Lambda)$ is obviously additive in M for fixed Λ. Since $Q(\mathrm{M}; \Lambda) \leqq P(\mathrm{M})$, $Q(\mathrm{M}; \Lambda)$ for fixed Λ is a completely additive function of sets M; hence integration with respect to the differential element $Q(de_1; \Lambda)$ has a meaning.

except possibly on a set Λ_ϵ such that $P(\Lambda_\epsilon) \leqq \epsilon/6$. If M is a P-measurable cylinder set over x_1, and if $a_j < \nu$, then

$$P(\mathrm{M}T^\nu\Lambda) = P(T^{a_j-\nu}\Lambda T^{a_j}\Lambda) = \int_{T^{a_j-\nu}\mathrm{M}} P(x_0; T^{a_j}\Lambda)dP, \tag{6.25}$$

so that, if use is made of (6.24) and (6.25),

$$\left| P(\mathrm{M}T^\nu\Lambda) - Q(T^{a_j-\nu}\mathrm{M}; \Lambda) \right| = \left| \int_{T^{a_j-\nu}\mathrm{M}} [P(x_0; T^{a_j}\Lambda) - Q(x_0; \Lambda)]dP \right|$$

$$\leqq \epsilon/6 + 2P(\Lambda_\epsilon) \leqq \epsilon/2,$$

if $\nu > a_j > N$. Then since

$$Q(T^k\mathrm{M}; \Lambda) = \lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} P(T^k\mathrm{M}T^j\Lambda) = \lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} P(\mathrm{M}T^{j-k}\Lambda)$$

is independent of k,

$$\lim_{\nu\to\infty} P(\mathrm{M}T^\nu\Lambda) = Q(\mathrm{M}; \Lambda), \tag{6.26}$$

so that (6.21) holds with $a_\nu = \nu$. The proof that (6.21) implies the existence of the limit in (6.23) can now be used to show that the existence of $\lim_{\nu\to\infty} P(\mathrm{M}T^\nu\Lambda)$ implies the existence of $\lim_{\nu\to\infty} P(x_0; T^\nu\Lambda)$. The hypothesis that F_x is the Borel field determined by a denumerable collection of its sets can now be removed; since if this is not true, we can preassign the set Λ, and then replace the field F_x by a smaller field F_x' for which the denumerability hypothesis is true, such that (cf. §2) the set Λ is in the corresponding field F_ω', and such that $\phi(x_0, x_1)$ is measurable with respect to F_ω'.

Conversely suppose that $\lim_{\nu\to\infty} P(x_0; T^\nu\Lambda)$ exists for every set Λ, as described in the theorem. Then if M is a P-measurable cylinder set over x_1,

$$\lim_{\nu\to\infty} P(\mathrm{M}T^\nu\Lambda) = \lim_{\nu\to\infty} \int_{\mathrm{M}} P(x_1; T^\nu\Lambda)dP \tag{6.27}$$

exists. Thus

$$\lim_{\nu\to\infty} \int f(x_1)\overline{g(x_\nu)}dP^* \tag{6.28}$$

exists, if $f(x_1)$, $g(x_1)$ are characteristic functions of cylinder sets of F_ω over x_1. The limit can then be shown to exist (using a familiar method of approximation) if $f(x_1)$, $g(x_1)$ are any bounded complex-valued P-measurable functions

* If ξ is a complex number, the notation $\bar{\xi}$ will be used, as is customary, to denote its conjugate complex number.

depending only on x_1. Now if there is an angle variable, there is, as was seen above, a bounded angle variable. If $\psi(x_1)$ is a bounded angle variable, we set $f=g=\psi$ in (6.28) and find that the limit

$$\lim_{\nu\to\infty}\int \psi(x_1)\overline{\psi(x_\nu)}dP = \lim_{\nu\to\infty} c^{-\nu+1}\int |\psi(x_1)|^2 dP, \qquad |c| = 1,\ c \neq 1,$$

must exist. This is absurd; hence there can be no angle variable.

Proof of (ii). If the process has no angle variables, and if it is also metrically transitive, $\lim_{\nu\to\infty} P(x_0; T^\nu\Lambda)$, which we know exists for Λ a P-measurable cylinder set over x_1 by part (i), must be $P(\Lambda)$ since (Theorem 6.1, Corollary 1)

$$\lim_{N\to\infty}\frac{1}{N}\sum_1^N P(x_0; T^m\Lambda) = P(\Lambda),$$

if we neglect x_0-sets of P-measure 0 throughout. Conversely if, whenever Λ is a P-measurable cylinder set over x_1, $\lim_{\nu\to\infty} P(x_0; T^\nu\Lambda) = P(\Lambda)$ except possibly on an x_0-set of P-measure 0, the process can have no angle variables, according to (i). If the process is not metrically transitive, there is a P-measurable cylinder set Λ over x_1, $(0<P(\Lambda)<1)$, which is invariant under T (if we neglect a set of P-measure 0). Then

$$P(x_0; \Lambda) = P(x_0; T^\nu\Lambda) \to P(\Lambda), \qquad \nu\to\infty,$$

that is, $P(x_0; \Lambda) = P(\Lambda)$ except possibly on an x_0-set of P-measure 0. This is incompatible with (6.11). The process therefore has no angle variables and is metrically transitive, as was to be proved.

Proof of (iii). We shall use the hypotheses of (iii) only to derive the fact that if Λ is a cylinder set of F_ω over x_1, and if $P(x_0; \Lambda) = 0$ for some value of x_0, then $P(\Lambda) = 0$. This fact is obvious from the equation

$$P(\Lambda) = \int P(x_0; \Lambda)P(dx_0).$$

Now consider the field of cylinder sets of F_ω over x_0, x_1. One probability measure is already defined on this field, namely P-measure. We define a second probability measure $\tilde{P}(\mathrm{M})$ for M in this field by

$$\tilde{P}(\mathrm{M}) = \int P(dx_0)\int f_{\mathrm{M}}(x_0, x_1)P(dx_1), \tag{6.29}$$

where $f_{\mathrm{M}}(x_0, x_1)$ is the characteristic function of M.* According to Theorem 3.4,

* This new measure is essentially a measure in two-dimensional (x_0, x_1)-space, obtained in the usual (multiplicative) way (cf. Saks, *Théorie de l'Intégrale*, Warsaw, 1933, pp. 257–263) from a given measure (P-measure) on the x_0-axis and a given measure (P-measure) on the x_1-axis.

$$P(\mathrm{M}) = \int P(de_0) \int f_{\mathrm{M}}(x_0, x_1) P(x_0; de_1), \tag{6.30}$$

and the integration need not be taken symbolically. Let $\mathrm{M}(x_0)$ be the cylinder set over x_1 defined by the equation $f_{\mathrm{M}}(x_0, x_1) = 1$. If we integrate in (6.30), then if $P(\mathrm{M}) = 0$, $P(x_0; \mathrm{M}(x_0)) = 0$, except possibly on an x_0-set of P-measure 0. It has already been shown that for each value of $x_0 = \xi$ such that $P(\xi; \mathrm{M}(\xi)) = 0$, $P(\mathrm{M}(\xi)) = 0$. Then if $P(\mathrm{M}) = 0$,

$$\tilde{P}(\mathrm{M}) = \int P(\mathrm{M}(x_0)) P(de_0) = 0;$$

hence the set function $\tilde{P}(\mathrm{M})$ is absolutely continuous with respect to $P(\mathrm{M})$. There is therefore* a function $\phi(x_0, x_1)$, measurable with respect to F_ω, such that if M is a cylinder set of F_ω over x_0, x_1,

$$\tilde{P}(\mathrm{M}) = \int_{\mathrm{M}} \phi(x_0, x_1) dP. \tag{6.31}$$

In particular if M is the intersection of Λ_0 (a cylinder set of F_ω determined by the condition $x_0 \in E_0$) and Λ (a cylinder set of F_ω determined by the condition $x_1 \in E_1$), (6.31) becomes

$$P(\Lambda_0) P(\Lambda) = \int_{E_0} P(de_0) \int_{E_1} P(x_0; de_1),$$

so that

$$\int_{E_0} P(de_0) \left\{ P(\Lambda) - \int_{E_1} \phi(x_0, x_1) P(x_0; de_1) \right\} = 0.$$

This equation is to hold for every set E_0 in the field F_x, so that the quantity in the brace must vanish, except possibly on an x_0-set of P-measure 0, as was to be proved.

THEOREM 6.6. *If the conditional probability functions of a temporally homogeneous stochastic process satisfy the conditions*

$$\begin{aligned} P(x_0; \Lambda) &\geqq \lambda_0 P(\Lambda) \\ P(x_{-\nu}, \cdots, x_0; \Lambda) &\geqq \lambda_\nu P(x_{-\nu+1}, \cdots, x_0; \Lambda), \quad \nu = 1, 2, \cdots, \end{aligned} \tag{6.32}$$

for every P-measurable cylinder set Λ over x_1,† where $0 < \lambda_\nu \leqq 1$, and if

* Saks, ibid., p. 257.

† The inequalities are to hold with probability 1 for each set Λ.

$$\prod_1^\infty \lambda_m^m = \lambda > 0, \tag{6.33}$$

then the process is metrically transitive and has no angle variables.

If the process is a Markoff process, we can take $\lambda_\nu = 1$, whenever $\nu > 0$, leaving only the first inequality of (6.32) as an actual condition. Theorem 6.2 gave a much more sensitive condition.

Let Λ_2 be a P-measurable cylinder set over $x_1, \cdots, x_n$, $n \geqq 1$. Then if f is the characteristic function of Λ_2, and if $m \geqq 1$ (cf. equation (3.10)), then

$$\begin{aligned} P(x_{-m}, \cdots, x_0; \Lambda_2) &= \int P(x_{-m}, \cdots, x_0; de_1) \int P(x_{-m}, \cdots, x_1; de_2) \int \\ &\quad \cdots \int fP(x_{-m}, \cdots, x_{n-1}; de_n) \\ &= 1 - \int P(x_{-m}, \cdots, x_0; de_1) \int \\ &\quad \cdots \int (1 - f)P(x_{-m}, \cdots, x_{n-1}; de_n) \\ &\leqq 1 - \int \lambda_m \lambda_{m-1} \cdots \lambda_0 P(de_1) \int \lambda_{m+1} \cdots \lambda_1 P(x_1; de_2) \int \\ &\quad \cdots \int \lambda_{m+n-1} \cdots \lambda_{n-1}(1 - f)P(x_1, \cdots, x_{n-1}; de_n) \\ &\leqq 1 - \lambda \int P(de_1) \int P(x_1; de_2) \int \\ &\quad \cdots \int (1 - f)P(x_1, \cdots, x_{n-1}; de_n) \\ &= 1 - \lambda[1 - P(\Lambda_2)]. \end{aligned} \tag{6.34}$$

If Λ_1 is a P-measurable cylinder set over $x_{-m}, \cdots, x_0$, $(m \geqq 0)$, then

$$P(\Lambda_1\Lambda_2) = \int_{\Lambda_1} P(x_{-m}, \cdots, x_0; \Lambda_2) dP \leqq P(\Lambda_1)\{1 - \lambda[1 - P(\Lambda_2)]\}. \tag{6.35}$$

Since this inequality is true for any sets Λ_1, Λ_2 as described, it is true for any P-measurable cylinder sets Λ_1, Λ_2 over $x_0, x_{-1}, \cdots; x_1, x_2, \cdots$ respectively. Now suppose there is a function $\phi(\omega)$, a complex-valued P-measurable function which does not vanish almost everywhere on Ω, and such that, for some constant c of modulus 1,

$$\phi(T\omega) = c\phi(\omega) \tag{6.36}$$

almost everywhere on Ω. To prove the theorem, it is sufficient to show that $\phi(\omega)$ is identically a constant almost everywhere on Ω. Since

$$\phi(T^\nu\omega) = c^\nu\phi(\omega), \qquad \nu = 1, 2, \cdots,$$

almost everywhere on Ω, if the integers $n_1, n_2, n_3, \cdots$ are chosen so that $\lim_{\nu\to\infty} c^{n_\nu} = 1$, it follows that

$$\lim_{\nu\to\infty} \phi(T^{n_\nu}\omega) = \lim_{\nu\to\infty} \phi(T^{-n_\nu}\omega) = \phi(\omega) \tag{6.37}$$

almost everywhere on Ω. Let $\Lambda = \Lambda(O)$ be the Ω-set defined by $\phi(\omega) \in O$ where O is an open set of the complex ϕ-plane. According to Lemma 6.1, (6.37) implies that $\Lambda(O)$ can be considered (neglecting sets of P-measure 0) as a cylinder set over both $x_1, x_2, \cdots$, and $x_{-1}, x_{-2}, \cdots$. Then in (6.35) we can take $\Lambda_1 = \Lambda_2 = \Lambda$, obtaining

$$P(\Lambda) \leqq P(\Lambda)\{1 - \lambda[1 - P(\Lambda)]\}, \tag{6.38}$$

which implies that $P(\Lambda) = 0$, or that $P(\Lambda) = 1$. Since O is arbitrary, this means that there is a constant ϕ_0 such that $\phi(\omega) = \phi_0$ almost everywhere,* as was to be proved.

As an application of the theorems of this section, we shall show how to derive a theorem of Kolmogoroff (I, p. 425).† Suppose that F_x is the Borel field determined by a denumerable collection of its sets,‡ and that conditional probability functions are given, as in Theorem 4.1 (ii), except that instead of supposing that (4.9) implies (4.10), we suppose, with Kolmogoroff, the validity of the stronger condition that there is a number $\lambda, (0 < \lambda \leqq 1)$, such that whenever Λ is a cylinder set of F_ω over x_1,

$$P(x_0; \Lambda) \geqq \lambda P(x_0'; \Lambda) \tag{6.39}$$

for all x_0, x_0'. There is then, according to Theorem 4.1 (ii) a temporally homogeneous Markoff process with the given conditional probability functions. From (6.36) (interchanging x_0, x_0') we find that

$$P(\Lambda) = \int P(x_0; \Lambda)P(de_0) \leqq \frac{1}{\lambda} P(x_0'; \Lambda) \tag{6.40}$$

for all x_0'. Then according to Theorem 6.6, with $\lambda_0 = \lambda, \lambda_1 = \lambda_2 = \cdots = 0$, the

* The point ϕ_0 of the complex plane is the intersection of the interiors of all the circles of rational radii with centers at points whose coordinates are rational and for which the corresponding Ω-sets are of P-measure 1.

† The proof to be given cannot compare in simplicity or elegance with that of Kolmogoroff. It is only given to show the significance of Kolmogoroff's hypothesis, (6.39) below, and the place of such a theorem in this development.

‡ This hypothesis will be eliminated below.

process is metrically transitive and has no angle variables. Moreover the hypotheses of Theorem 6.5 (iii) are satisfied, so that, according to part (ii) of that theorem, $P(x_0; T^m\Lambda) \to P(\Lambda)$ except possibly on an x_0-set of P-measure 0. We shall show that this exceptional set is actually empty. In the integral following

$$\int P(x_1; T^m\Lambda)P(x_0; de_1) = P(x_0; T^m\Lambda) \tag{6.41}$$

we have just seen that the integrand converges to $P(\Lambda)$ except possibly on an x_1-set of P-measure 0. It follows readily from (6.39) that if $P(\Lambda)=0$, then $P(x_0; \Lambda)\equiv 0$,* so that the exceptional set is of $P(x_0; de_1)$-measure 0 for each value of x_0. Then term by term integration in (6.41) gives Kolmogoroff's result, that $P(x_0; T^m\Lambda) \to P(\Lambda)$ for all x_0.† The assumption made above, that F_x is the Borel field determined by a denumerable collection of its sets, is unnecessary, since in any case, if Λ is preassigned, F_x can be chosen to satisfy the denumerability condition and the condition that Λ lies in F_ω.

7. **Application to the examples of** §5. In this section we apply the results of §6 to a detailed study of the examples of §5.

I. In this case, that of a sequence of mutually independent chance variables, the conditional probabilities become absolute probabilities. If the process is temporally homogeneous, it is always metrically transitive and has no angle variables (Theorem 6.4). The ergodic theorem, as applied in Theorem 6.1, gives the strong law of large numbers.‡

II. We have seen above (§5) that absolute probabilities $p_1, \cdots, p_n$ always exist in case II, and can be obtained in the form

$$p_k = \lim_{\nu\to\infty} \frac{1}{N_\nu} \sum_{m=1}^{N_\nu} p_{jk}^{(m)}, \qquad (j \text{ fixed}). \tag{7.1}$$

THEOREM 7.1. (i) *Except possibly on an* Ω*-set of* P*-measure* 0,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{m=1}^{N} p_{x_m k}. \tag{7.2}$$

exists.§

* In fact $P(\Lambda) \geqq \lambda P(x_0'; \Lambda)$ for all x_0'.

† Kolmogoroff actually obtains more, since he obtains an estimate of the speed of convergence.

‡ Cf. Doob (I, pp. 764–765); Hopf (I, p. 83); Khintchine (I).

§ Part (i) supposes that some set of absolute probabilities is accepted, thus determining P-measure. In (7.2), $p_{x_m k}$ is a chance variable, a function of ω: $(\cdots, x_{-1}, x_0, \cdots)$ taking on the value p_{rk} at ω if $x_m = r$. Since only cylinder sets over $x_1, x_2, \cdots$ are involved in the theorem, the result holds when only the space of points $(x_1, x_2, \cdots)$ (on which P-measure is defined in terms of that on Ω in an obvious way) is considered.

(ii) *If (j, k) is any pair of subscripts, there exists*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{m=1}^{N} p_{jk}^{(m)} = q_{jk}. \tag{7.3}$$

The existence of the limit in (7.3) was proved by Fréchet (I, p. 151) by means of an explicit determination of $p_{jk}^{(m)}$, as a function of j, k, m, derived from the theory of linear difference equations. The proof given here will hold in the more general case III. It will be remembered that (7.3) is an integrated form of (7.2).

Let $p_1, \cdots, p_n$ be some set of absolute probabilities corresponding to the given matrix. The first part of the theorem is simply the first part of Theorem 6.1 (cf. equation (6.2)) in this special case. The proof of (ii) requires more care. We shall first choose a particular set of absolute probabilities $p_1, \cdots, p_n$ obtained by applying the corollary of Theorem 4.1 (ii), where we define the set function $Q(E)$ to be equal to the number of points in E divided by n. This convention gives absolute probabilities defined by

$$p_k = \frac{1}{n} \lim_{\nu\to\infty} \frac{1}{N_\nu} \sum_{m=1}^{N_\nu} \left(\sum_{j=1}^{n} p_{jk}^{(m)} \right). \tag{7.4}$$

According to Theorem 6.1, the limit in (7.3) exists for all pairs of subscripts j, k for which $p_j > 0$. Let J be the set of subscripts j for which $p_j = 0$. Then the limit in (7.3) exists if $j \in J$. We can write $p_{jk}^{(m+\nu)}$ in the following form:

$$p_{jk}^{(m+\nu)} = \sum_{l=1}^{n} p_{jl}^{(\nu)} p_{lk}^{(m)}, \tag{7.5}$$

and if we set

$$\Pi_{jk}^{(N)} = \frac{1}{N} \sum_{m=1}^{N} p_{jk}^{(m)},$$

we obtain the relation

$$\frac{\nu+\mu}{\mu} \Pi_{jk}^{(\nu+\mu)} - \frac{\nu}{\mu} \Pi_{jk}^{(\nu)} = \sum_{l=1}^{n} p_{jl}^{(\nu)} \Pi_{lk}^{(\mu)}. \tag{7.6}$$

Now according to (7.4), since $p_l = 0$ if $l \in J$,

$$\lim_{\nu\to\infty} \frac{1}{N_\nu} \sum_{m=1}^{N_\nu} \left[\sum_{j=1}^{n} \left(\sum_{l\in J} p_{jl}^{(m)} \right) \right] = 0. \tag{7.7}$$

This implies that

$$\liminf_{m\to\infty} \sum_{j=1}^{n} \sum_{l\in J} p_{jl}^{(m)} = 0. \tag{7.8}$$

Moreover, it has already been shown that

$$\lim_{\mu\to\infty} \Pi_{lk}^{(\mu)} = q_{lk}, \qquad l \notin J, \tag{7.9}$$

exists. Then letting μ become infinite in (7.6), and using (7.9), we obtain

$$\limsup_{N\to\infty} \Pi_{jk}^{(N)} = \limsup_{\mu\to\infty} \sum_{l\in J} p_{jl}^{(\nu)} \Pi_{lk}^{(\mu)} + \sum_{l\notin J} p_{jl}^{(\nu)} q_{lk} \leqq \sum_{l\in J} p_{jl}^{(\nu)} + \sum_{l\notin J} p_{jl}^{(\nu)} q_{lk} \tag{7.10}$$

and

$$\liminf_{N\to\infty} \Pi_{jk}^{(N)} \geqq \sum_{l\notin J} p_{jl}^{(\nu)} q_{lk}, \tag{7.11}$$

so that

$$\limsup_{N\to\infty} \Pi_{jk}^{(N)} - \liminf_{N\to\infty} \Pi_{jk}^{(N)} \leqq \sum_{l\in J} p_{jl}^{(\nu)}. \tag{7.12}$$

This inequality is true for $\nu = 1, 2, \cdots$, so that, using (7.8), we obtain

$$\limsup_{N\to\infty} \Pi_{jk}^{(N)} = \liminf_{N\to\infty} \Pi_{jk}^{(N)} \tag{7.13}$$

as was to be proved.

Since, in general, there is not a unique set of absolute probabilities $p_1, \cdots, p_n$, a given matrix may correspond to several temporally homogeneous processes. If all these processes are metrically transitive, the matrix (p_{jk}) will be called metrically transitive. If none of these processes has angle variables, the matrix will be said to have no angle variables. Otherwise the matrix will be said to be not metrically transitive, or to have angle variables, as the case may be.

THEOREM 7.2. *The matrix (p_{jk}) is metrically transitive if and only if*

(i) *there is a single set of absolute probabilities $(p_1, \cdots, p_n)$; or*

(ii) *the limit q_{jk} depends only on k; or*

(iii) *the equations*

$$\sum_{j=1}^{n} x_j p_{jk} = x_k, \qquad k = 1, \cdots, n, \tag{7.14}$$

have only a single linearly independent solution in $(x_1, \cdots, x_n)$, that is, the matrix $(p_{jk} - \delta_{jk})$ has rank $n-1$; or*

* This condition is not the same as that of (i) since the absolute probabilities are restricted to be non-negative.

(iv) *the characteristic equation of the matrix* (p_{ik}) *has* 1 *as a simple root; or*

(v) *the matrix* (p_{ik}) *cannot be put in the form of Fig.* 1 (*where* R_1, R_2, R_3 *are square matrices and the* 0*'s represent blocks consisting entirely of* 0*-elements, in which* R_3, *but not* R_1 *or* R_2, *may be absent, by means of some permutation applied to both rows and columns.*

$$\begin{pmatrix} R_1 & 0 & 0 \\ 0 & R_2 & 0 \\ \cdot & \cdot & R_3 \end{pmatrix}$$

Fig. 1

It would be very difficult to give complete references to previous work on the various parts of this and the following theorems, and such references are perhaps made unnecessary by Fréchet's forthcoming book. Since the time of Markoff, various writers have rediscovered and extended his results, independently of Markoff and of each other. It is hoped that this paper will provide a certain unity to these results, and it is claimed that the terminology used to describe the various cases is of more general validity and less *ad hoc* than that previously used. The methods, and some of the results, are new. Fréchet and Hadamard (I) have given a historical discussion of some of them. The equivalence of (i)–(iv) was shown by Fréchet (I) in the most detailed treatment of case II which has as yet appeared. The equivalence of (ii) and (v) is somewhat related to more specialized results of von Mises (I, pp. 533–549). The equivalence of (iv) and (v) (in a somewhat different form, with the additional hypothesis that no column of (p_{ik}) contains only 0 elements) was obtained by Romanovsky (I, pp. 154–155) by applying theorems of Frobenius. The matrix can be further decomposed if 1 is a root of multiplicity >2. As Romanovsky proves, and as follows readily here also, R_1, R_2 can be replaced by ν boxes along the main diagonal, if ν is the multiplicity of 1 as a root of the characteristic equation. A complete proof of each part of Theorem 7.2 will be given, since the method will be available for the treatment of case III, and the details of the latter case will then be omitted.

Proof of (i). Suppose that the given matrix is metrically transitive, and let $p_1, \cdots, p_n$ be a set of absolute probabilities corresponding to it. Then according to Theorem 6.1, Corollary 1,

$$q_{ik} = p_k, \qquad k = 1, \cdots, n,$$

if $p_i > 0$. If $p_1', \cdots, p_n'$ is a second set of absolute probabilities corresponding to the given matrix, then

$$\tfrac{1}{2}(p_1 + p_1'), \cdots, \tfrac{1}{2}(p_n + p_n')$$

is also a set of absolute probabilities corresponding to the given matrix, so that if $p_j+p_j' > 0$,

$$q_{jk} = \tfrac{1}{2}(p_k + p_k'), \qquad k = 1, \cdots, n.$$

Combining these two results, if we choose j_0 so that $p_{j_0} > 0$,

$$q_{j_0 k} = p_k = \tfrac{1}{2}(p_k + p_k'), \qquad k = 1, \cdots, n,$$

that is,

$$p_k = p_k', \qquad k = 1, \cdots, n.$$

Thus metric transitivity implies that there is only a single set of absolute probabilities.

Conversely, suppose that there is only a single set of absolute probabilities, $p_1, \cdots, p_n$. It can be verified directly that for each value of j, $q_{j1}, \cdots, q_{jn}$ is a set of absolute probabilities corresponding to the given matrix,* so that $q_{jk} = p_k$ for all j, k. If the matrix is not metrically transitive, the process determined by the matrix of conditional probabilities (p_{jk}) and the absolute probabilities is not metrically transitive (that is, the corresponding transformation T is not metrically transitive), so that there is, according to Theorem 6.2, a set of subscripts K, such that

$$0 < p_k, \quad k \in K, \qquad \sum_{k \in K} p_k < 1, \tag{7.15}$$

$$\sum_{l \in K} p_{kl}^{(m)} = 1, \qquad k \in K, \qquad m = 1, 2, \cdots.\dagger \tag{7.16}$$

Then

$$\sum_{l \in K} \Pi_{kl}^{(m)} = 1, \qquad k \in K, \qquad m = 1, 2, \cdots,$$

so that, using the fact that $\Pi_{kl}^{(m)} \to q_{kl} = p_l$, and (7.16), we obtain

$$\sum_{l \in K} p_l = \lim_{m \to \infty} \sum_{l \in K} \Pi_{kl}^{(m)} = 1, \qquad k \in K,$$

contradicting (7.15). The matrix is thus metrically transitive.

---

* In general, it can be shown that if $q_1, \cdots, q_n$ is a linear combination of columns of the matrix (q_{ij}), where the coefficients of the combination are non-negative and have sum 1, then $q_1, \cdots, q_n$ is a set of absolute probabilities corresponding to the given matrix, and conversely every set of absolute probabilities corresponding to the given matrix can be obtained in this way. Cf. the discussion of case III below.

† The Ω-set determined by the condition $x_0 \in K$ is invariant under T up to an x_0-set of P-measure 0. Equation (7.16) for $m=1$ is then the first equation of (6.11), and it follows for $m>1$ by direct verification in view of the definition of $p_{jk}^{(m)}$.

Proof of (ii). If the matrix is metrically transitive, it was shown in the proof of (i) that $q_{jk}=p_k$ for all j, k (where $p_1, \cdots, p_n$ is the uniquely determined set of absolute probabilities) and q_{jk} therefore depends only on k. Conversely if $q_{jk}=q_k$ is independent of j, we shall show that the absolute probabilities are uniquely determined by investigating the solutions of (7.14). Let $(x_1, \cdots, x_n)$ be a solution of (7.14). Then it can be verified directly that

$$\sum_{j=1}^{n} x_j p_{jk}^{(m)} = x_k, \qquad k = 1, \cdots, n; \; m = 1, 2, \cdots, \tag{7.17}$$

so that

$$\sum_{j=1}^{n} x_j \Pi_{jk}^{(N)} = x_k, \qquad k = 1, \cdots, n; \; N = 1, 2, \cdots. \tag{7.18}$$

Letting m become infinite in (7.18) we obtain

$$\sum_{j=1}^{n} x_j q_k = q_k \sum_{j=1}^{n} x_j = x_k, \qquad k = 1, \cdots, n. \tag{7.19}$$

Then if $(p_1, \cdots, p_n)$ is a set of absolute probabilities, since $p_1, \cdots, p_n$ is a solution of (7.14) and since $p_1+\cdots+p_n=1$,

$$q_k \sum_{j=1}^{n} p_j = q_k = p_k, \qquad k = 1, \cdots, n.$$

Thus the absolute probabilities are uniquely determined; which fact implies, according to (i) that the matrix (p_{jk}) is metrically transitive.

Proof of (iii). If the system (7.14) has only a single linearly independent solution, the absolute probabilities (which constitute a particular solution) are surely uniquely determined, so the matrix is metrically transitive (according to (i)). Conversely, according to (i), if the matrix is metrically transitive, there is a unique set of absolute probabilities $p_1, \cdots, p_n$, and we have seen that $q_{jk}=p_k=q_k$, $(j, k=1, \cdots, n)$. Then if $(x_1, \cdots, x_n)$ is any solution of (7.14), it is linearly dependent on $(p_1, \cdots, p_n)$ by (7.19).

Proof of (iv). Since a set of absolute probabilities is a solution of (7.14), 1 is always a root of the characteristic equation of the matrix (p_{jk}). If the matrix is metrically transitive, there is only a single linearly independent solution of (7.14), according to (iii), and we shall show that 1 is a simple root of the characteristic equation of the matrix. Suppose the contrary. If all the columns of the matrix $(p_{jk}-\lambda\delta_{jk})$ are added to the first, every element of the first column becomes $1-\lambda$. If $1-\lambda$ is factored from the determinant, the determinant still vanishes for $\lambda=1$, by hypothesis. Then the equations

$$x_1 + \cdots + x_n = 0,$$

$$\sum_{j=1}^{n} x_j p_{jk} = x_k, \qquad k = 2, \cdots, n, \tag{7.20}$$

have a non-trivial solution (expressing the fact that the rows of the determinant are linearly dependent). If the last $n-1$ equations are subtracted from the first, the first becomes

$$\sum_{j=1}^{n} x_j p_{j1} = -x_2 - \cdots - x_n = x_1. \tag{7.21}$$

Thus $(x_1, \cdots, x_n)$ is a solution of (7.14), and since $x_1 + \cdots + x_n = 0$, it is not linearly dependent on $(p_1, \cdots, p_n)$; which contradicts the fact that there is only a single linearly independent solution of (7.14).

Conversely, if 1 is a simple root of the characteristic equation of the matrix (p_{jk}), the system (7.14) has only a single linearly independent solution.*

Proof of (v). If the given matrix is metrically transitive, we shall prove, using the fact that $q_{jk} = p_k$ for all j, k (where $p_1, \cdots, p_n$ is the uniquely determined set of absolute probabilities), that the matrix (p_{jk}) cannot be put in the form of Fig. 1, with R_1 and R_2 both present, by a transformation of the type described. (Such a transformation corresponds to a relabeling of the points of X.) If, on the contrary, the matrix can be put in this form, it is no restriction to assume that it is already in this form. It can then be verified directly that the iterated matrix $(p_{jk}^{(m)})$, and therefore $\Pi_{jk}^{(N)}$, will also be in this form with the same blocks R_1, R_2. But then each column of the limiting matrix $(q_{jk}) = (p_k a_j)$ (with $a_1 = \cdots = a_n = 1$) contains zeros, so that $p_1 = \cdots = p_n = 0$, contrary to fact.

To show the converse we shall assume that the matrix is not metrically transitive and put it in the form described. Let $p_1, \cdots, p_n$ be a set of absolute probabilities for which the corresponding process is not metrically transitive. If any p's vanish, we can assume they are the last ones:†

$$p_j > 0, \qquad j \leqq \alpha, \qquad 0 < \alpha \leqq n,$$
$$p_{\alpha+1} = p_{\alpha+2} = \cdots = p_n = 0. \tag{7.22}$$

Since

$$\sum_{j=1}^{n} p_j p_{jk} = p_k, \qquad k = 1, \cdots, n,$$

* This follows from elementary matrix theory and does not depend upon the particular properties of our matrix (p_{jk}).

† This assumption implies the possibility of a matrix transformation of the type described in the theorem.

it follows that

$$p_{jk} = 0, \quad \text{if} \quad j \leqq \alpha, \qquad k > \alpha, \tag{7.23}$$

where the inequalities on j, k are to hold simultaneously. Because of the fact that there is not metric transitivity, there is a set of subscripts K such that $p_j > 0$ if $j \in K$, that $\sum_{j \in K} p_j < 1$, and that

$$p_{jk} = 0, \quad \begin{cases} j \in K, & k \notin K \\ j \notin K, & k \in K, \quad j \leqq \alpha. \end{cases}^{*} \tag{7.24}$$

We can assume that K consists of the first a subscripts; then equations (7.23) and (7.24) describe the form of Fig. 1. Since K is not empty, R_1 cannot be absent; since $\sum_{j \in K} p_j < 1$, R_2 cannot be absent. If no p_j vanishes, R_3 is absent.

THEOREM 7.3. *The matrix (p_{jk}) has no angle variables if and only if*

(i) *for every pair of subscripts j, k,*

$$\lim_{m \to \infty} p_{jk}^{(m)} = q_{jk} \tag{7.25}$$

exists; or

(ii) *1 is the only root of modulus 1 of the characteristic equation of p_{jk}; or*

(iii) *the matrix cannot be put in the form of Fig. 1 by a transformation of the type described in Theorem 7.2 (v), where R_2, R_3 may not be present, and where R_1 is itself in the form of Fig. 2, obtained by dividing the subscripts into ν groups $J_1, \cdots, J_\nu$, of consecutive subscripts, such that $p_{jk} = 0$, unless p_{jk}, is in some one of the square matrices $S_1, \cdots, S_\nu$ for which $j \in J_r$, $k \in J_{r+1}$, $(r < \nu)$,* or *$j \in j_\nu$, $k \in J_1$.*†

$$\begin{array}{c} \\ J_1 \\ J_2 \\ J_3 \\ J_4 \end{array} \begin{array}{c} \begin{array}{cccc} J_1 & J_2 & J_3 & J_4 \end{array} \\ \begin{pmatrix} 0 & S_1 & 0 & 0 \\ 0 & 0 & S_2 & 0 \\ 0 & 0 & 0 & S_3 \\ S_4 & 0 & 0 & 0 \end{pmatrix} \end{array}, \qquad \nu = 4,$$

Fig. 2

Proof of (i). Suppose that the matrix (p_{jk}) has no angle variables. Let

* Cf. Theorem 6.2. The Ω-set determined by the condition $x_0 \in K$ is invariant under T, if we neglect a set of P-measure 0. The first set of equations in (7.24) is obtained from the first equation of (6.11), which implies that $P(x_0; C\Lambda) = 0$, $(x_0 \in E)$, except perhaps for an x_0-set of P-measure 0, and the second set of equations in (7.24) is obtained directly from the second equation of (6.11).

† The equivalence of (i), (ii) was shown by Fréchet (I, q.v. for earlier references). Romanovsky (I) has discussed matrices like that of Fig. 2. Cf. also Doeblin and Fortet (I, p. 1700). The latter authors omit mention of exceptional points, implying here the possible existence of R_3.

$p_1, \cdots, p_n$ be a set of absolute probabilities corresponding to this matrix. The condition of Theorem 6.5 will be satisfied if, whenever r is a subscript such that $p_r>0$, p_{rk} can be put in the form

$$p_{rk} = \phi_{rk}p_k, \qquad k = 1, \cdots, n.$$

This will be possible if, for r fixed, $p_{rk}=0$ whenever $p_k=0$; but this is true, since

$$\sum_{j=1}^{n} p_j p_{jk} = p_k.$$

Thus the condition of Theorem 6.5 is satisfied so that the limit in (7.25) exists for all j, k for which $p_j>0$. Then if J is the set of subscripts for which the absolute probabilities vanish, the limit in (7.25) exists if $j \notin J$. We shall suppose, using the results of Theorem 7.1 and the fact that for any subscript j $q_{j1}, \cdots, q_{jn}$ is a set of absolute probabilities corresponding to the given matrix, that the absolute probabilities $p_1, \cdots, p_n$ are given by

$$p_k = \frac{1}{n}\sum_{j=1}^{n} q_{jk} = \frac{1}{n}\lim_{N\to\infty}\frac{1}{N}\sum_{m=1}^{N}\sum_{j=1}^{n} p_{jk}^{(m)}. \tag{7.26}$$

We can write $p_{jk}^{(\mu+\nu)}$ in the form

$$p_{jk}^{(\mu+\nu)} = \sum_{l\in J} p_{jl}^{(\mu)} p_{lk}^{(\nu)} + \sum_{l\notin J} p_{jl}^{(\mu)} p_{lk}^{(\nu)}. \tag{7.27}$$

Then letting ν become infinite in (7.27), we obtain

$$\begin{aligned} \limsup_{m\to\infty} p_{jk}^{(m)} &\leqq \sum_{l\in J} p_{jl}^{(\mu)} + \sum_{l\notin J} p_{jl}^{(\mu)} q_{lk} \\ \liminf_{m\to\infty} p_{jk}^{(m)} &\geqq \sum_{l\notin J} p_{jl}^{(\mu)} q_{lk}, \end{aligned} \tag{7.28}$$

so that

$$\limsup_{m\to\infty} p_{jk}^{(m)} - \liminf_{m\to\infty} p_{jk}^{(m)} \leqq \sum_{l\in J} p_{jl}^{(\mu)} \qquad \mu = 1, 2, \cdots. \tag{7.29}$$

Now as in the proof of Theorem 7.1 we know that (7.8) is true, and combining this with (7.29) we obtain

$$\limsup_{m\to\infty} p_{jk}^{(m)} = \liminf_{m\to\infty} p_{jk}^{(m)}$$

as was to be proved.

Conversely, if the limit in (7.25) exists for every pair of subscripts j, k, Theorem 6.5 shows that there can be no angle variables in any process corresponding to the matrix (p_{jk}); thus the matrix (p_{jk}) has no angle variables.

Proof of (ii). If there are no angle variables corresponding to the given matrix, there can be no root of the characteristic equation of the matrix (p_{jk}) of modulus 1, other than 1; for if c is such a root, there is a set of constants $x_1, \cdots, x_n$, not all 0, such that

$$\sum_{k=1}^{n} p_{jk}x_k = cx_j, \qquad j = 1, \cdots, n. \tag{7.30}$$

Then

$$\sum_{k=1}^{n} p_{jk}^{(m)} x_k = c^m x_j, \quad j = 1, \cdots, n; m = 1, 2, \cdots. \tag{7.31}$$

When m becomes infinite, the left side converges, to $\sum_{k=1} q_{jk}x_k$, whereas the right side, if j is chosen so that $x_j \neq 0$, does not converge. Then the characteristic root c is impossible.

Conversely, suppose that there is no root of the characteristic equation of modulus 1 other than 1. Let $p_1, \cdots, p_n$ be any set of absolute probabilities corresponding to the given matrix. We shall prove that the temporally homogeneous process defined in terms of $p_1, \cdots, p_n$ and (p_{jk}) can have no angle variable, by showing that the existence of an angle variable implies the existence of a root (not equal to 1) of the characteristic equation, of modulus 1. If $\psi(x_0)$ is an angle variable, and if $\psi(j) = \psi_j$, (6.18) becomes

$$\sum_{k=1}^{n} p_{jk}\psi_k = c\psi_j, \qquad |c| = 1, c \neq 1, \tag{7.32}$$

for those values of j for which $p_j > 0$. Let these make up the set of subscripts not belonging to the set J. If $j \notin J$, we have seen above that $p_{jk} = 0$ for $k \in J$, so that changing ψ_k for $k \in J$ does not affect (7.32). We shall attempt to re-define ψ_j for $j \in J$ to make (7.32) valid for all j. To do this, we must solve the following system of equations for $\psi_j, j \in J$:

$$\sum_{k=1}^{n} p_{jk}\psi_k = c\psi_j, \qquad j \in J; \tag{7.33}$$

or, if we set

$$\sum_{k \notin J} p_{jk}\psi_k = \alpha_j,$$

the system

$$\sum_{k \in J} p_{jk}\psi_k = c\psi_j - \alpha_j, \qquad j \in J, \tag{7.34}$$

that is,

$$\sum_{k \in J} (p_{jk} - c\delta_{jk})\psi_k = -\alpha_j, \qquad j \in J. \tag{7.35}$$

If these equations have a solution, this solution, when combined with the ψ_j's for $j \notin J$, satisfies (7.32) for all j, so that the matrix (p_{jk}) has the number c as a root of its characteristic equation. On the other hand, if the system (7.35) has no solution, the matrix (p_{jk}), with j, k restricted to J, has c as a root of its characteristic equation, so that there is a set of numbers $\{\gamma_j\}$, $j \in J$, not all 0, satisfying

$$\sum_{k \in J} p_{jk}\gamma_k = c\gamma_j, \qquad j \in J. \tag{7.36}$$

But then if γ_j is defined as 0 for $j \notin J$, the set $\gamma_1, \cdots, \gamma_n$ provides a non-trivial solution of (7.32) for all j, so that again the matrix (p_{jk}) has the number c as a root of its characteristic equation. In any case then, the hypothesis that there is an angle variable implies the existence of a root $c \neq 1$, $|c| = 1$, (which is the characteristic value corresponding to the angle variable) of the characteristic equation of the matrix.

Proof of (iii). If there are no angle variables, the matrix (p_{jk}) cannot be put in the form described; for it is readily verified that if (p_{jk}) is in this form, the matrices $(p_{jk}^{(\nu+1)})$, $(p^{(2\nu+1)})$, $\cdots$ are of the same form, whereas the matrices $(p_{jk}^{(2)})$, $(p_{jk}^{(\nu+2)})$, $(p_{jk}^{(2\nu+2)})$, $\cdots$ are of the same form except that the non-zero blocks of R_1 are the matrices determined by (J_1J_3), (J_2J_4), $\cdots$ instead of (J_1J_2), (J_2J_3), $\cdots$. Then if $p_{jk}^{(m)} \to q_{jk}$, the submatrix R_1 of (q_{jk}) must have only 0 elements; but this contradicts the fact that the sum of the elements in each row of R_1 is 1. (It would also have been possible to prove this part by giving an explicit definition of an angle variable corresponding to the given matrix.)

Conversely, suppose that there is an angle variable corresponding to some choice $p_1, \cdots, p_n$ of the absolute probabilities, so that (Theorem 6.3 (i)) there is a function $\psi(x_0)$ such that

$$\psi(x_1) = c\psi(x_0), \qquad |c| = 1, c \neq 1, \tag{7.37}$$

except on an (x_0, x_1)-set of P-measure 0. Since $\psi(x_0)$ (which takes on at most n values) necessarily takes on some value on a set of positive P-measure, c must be a root of unity (Theorem 6.3 (ii)). This fact will appear again below. Let $\psi(j) = \psi_j$. Let $a_1, a_2, \cdots$ be those non-zero values in the set $\psi_1, \cdots, \psi_n$

for which the corresponding probability p_j is positive. Define J_k as the set of subscripts j for which $p_j>0$, and $\psi_j=a_k$. Let n_k be the number of subscripts in J_k. We can assume (transforming the matrix as described above if necessary) that J_1 consists of the first n_1 subscripts, J_2 of the next n_2 subscripts, and so on. According to (7.37), some a_j will necessarily be ca_1, and we can suppose it to be a_2. In the same way, some a_j will necessarily be ca_2, and we can assume it to be a_3 (unless it is a_1), $\cdots$. Continuing this, we will necessarily find a first integer $\nu>1$, such that if $a_1, a_2, \cdots, a_r$ are chosen successively as described, so that $a_2=ca_1, \cdots, a_r=c^{r-1}a_1$, the next application of the algorithm will give $ca_r=a_1$, and hence $c^r=1$. Then c is a νth root of unity, and $1<\nu\leqq n$. If $n_{r-1}<x_0\leqq n_r$* (so that $\psi(x_0)=a_r$), then if $r<\nu$, it follows that $\psi(x_1)=a_{r+1}$ necessarily (if $r=\nu$, $\psi(x_1)=a_1$ necessarily), if we neglect x_1-sets of P-measure 0, that is, subscripts k for which $p_k=0$; and

$$p_{jk}=0 \quad \text{if} \quad \begin{cases} j\in J_r, & k\notin J_{r+1}, \quad p_k>0, \\ j\in J_\nu, & k\notin J_1, \quad p_k>0. \end{cases} \qquad r=1,\cdots,\nu-1,$$

These equations describe the $(J_1+\cdots+J_\nu)^2$ matrix R_1. The fact that the Ω-set determined by the condition $x_1\in J_1+\cdots+J_\nu$ is invariant under the transformation T up to a set of P-measure 0 means that the matrix (p_{jk}) can be put in the form of Fig. 1, as was shown in the proof of the preceding theorem, except that in this case the matrices R_2, R_3 may be absent. The R_3 is absent if every p_j is positive; R_2 is absent if the Ω-set, determined by the condition $x_0\in J_1+\cdots+J_\nu$, has P-measure 1.

THEOREM 7.4. *The matrix (p_{jk}) is metrically transitive and has no angle variables if and only if*

(i) *for every pair of subscripts j, k, $\lim_{m\to\infty} p_{jk}^{(m)}$ exists and is independent of j; or*

(ii) *the root 1 is the only root of the characteristic equation of modulus 1 and is itself a simple root; or*

(iii) *the matrix cannot be put in the form of Fig. 1 if either both R_1 and R_2 are present, or if R_1 is present and has the form of Fig. 2; or*

(iv) *each matrix $(p_{jk}^{(1)}), \cdots, (p_{jk}^{(n)})$ is metrically transitive.*

Only the last part requires any comment. Suppose that the matrix (p_{jk}) is metrically transitive and has no angle variables. Then $p_{jk}^{(m)}\to p_k$ (where $p_1, \cdots, p_n$ is the uniquely determined set of absolute probabilities). In particular, $\lim_{m\to\infty} p_{jk}^{(rm)}=p_k$, so that, according to (i) the matrix $(p_{jk}^{(r)})$ is metrically transitive. Conversely, suppose that the matrices $(p_{jk}^{(1)}), \cdots, (p_{jk}^{(n)})$ are

* Take $n_0=0$.

metrically transitive. We shall prove that the matrix (p_{jk}) can have no angle variable. If there is an angle variable, its characteristic value c, $(|c| = 1, c \neq 1)$, is a root of the characteristic equation of the matrix (p_{jk}) (cf. the proof of Theorem 7.3 (iii)), and we have seen that this number c must be a root of unity of order less than or equal to n. Then there is a set of numbers $x_1, \cdots, x_n$ (not all 0) such that

$$\sum_{j=1}^{n} x_j p_{jk} = c x_k, \qquad k = 1, \cdots, n.$$

Moreover, if we sum over k,

$$\sum_{j=1}^{n} x_j = c \sum_{k=1}^{n} x_k,$$

which implies that $\sum_{j=1}^{n} x_j = 0$. If ν is chosen so that $c^\nu = 1$, $(\nu \leqq n)$, then $\sum_{j=1}^{n} x_j p_{jk}^{(\nu)} = c^\nu x_k = x_k$. Now the (uniquely defined) absolute probabilities $p_1, \cdots, p_n$ satisfy the equations

$$\sum_{j=1}^{n} p_j p_{jk}^{(\nu)} = p_k, \qquad k = 1, \cdots, n.$$

If the matrix $(p_{jk}^{(\nu)})$ is to be metrically transitive, the sets $p_1, \cdots, p_n$, $x_1, \cdots, x_n$ must be linearly dependent (Theorem 7.2 (iii)), but this is impossible, since

$$\sum_{j=1}^{n} p_j = 1, \qquad \sum_{j=1}^{n} x_j = 0.$$

The hypothesis that the matrix (p_{jk}) has an angle variable has thus led to a contradiction.

III. In this example, special conditions must be imposed on the conditional probability density $p(x, y)$ to insure the existence of an absolute probability function. If there is an absolute probability function, it was shown in §5 that it is determined by an X-integrable density function $p(x)$ which can be supposed to satisfy

$$\int p(x)dx = 1, \qquad \int p(x)p(x, y)dx = p(y) \tag{7.38}$$

for all $y \in X$. In the following we shall mean by an absolute probability density $p(x)$ an X-integrable function, which is non-negative and satisfies (7.38) for all y. It follows that two absolute probability densities which are equal almost everywhere on X are identical. *We shall assume throughout that* $p^{(m)}(x, y)$, *for*

some integer $m \geqq 1$, satisfies the condition of uniform integrability discussed in §5 *insuring the existence of at least one absolute probability density function.* In fact, we have seen in the corollary to Theorem 4.1 that if $Q(E)$ is any probability measure defined on the sets of F_x, we can obtain a probability density $p(x)$ in the form

$$\int_E p(x)dx = \lim_{\nu\to\infty} \frac{1}{N_\nu} \sum_{m=1}^{N_\nu} \int Q(dx) \int_E p^{(m)}(x, y)dy. \tag{7.39}$$

The denumerability hypothesis on the field F_x and the uniform integrability hypothesis were employed in order to obtain a certain compactness in an aggregate of set functions (cf. §4) through which (7.39) was derived. More abstract formulations are possible (cf. the hypotheses of Kryloff and Bogolioùboff (I)).

THEOREM 7.5. (i) *Except possibly on an Ω-set of P-measure* 0,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{m=1}^{N} p(x_m, y) = q(\omega; y)^* \tag{7.40}$$

exists for each value of y for which $p(y)$ is finite-valued.

(ii) *The limit*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{m=1}^{N} \int_E p^{(m)}(x, y)dy = q(x, E) \tag{7.41}$$

exists for all $x \in X$ and every set $E \in F_x$. If there is a value of m for which $p^{(m)}(x, y)$ is a bounded function of x for each value of y,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{m=1}^{N} p^{(m)}(x, y) = q(x, y) \tag{7.42}$$

exists for all x, y.

Part (i) of the theorem is new. Part (ii), which is an integrated form of (i), generalizes Theorem 7.1 (ii). The existence of the limit in (7.42) was proved by Fréchet (II, p. 81) who supposes that X is the closed cover of a bounded domain of euclidean space, and that there is a value of m such that $p^{(m)}(x, y)$ is a bounded function. Fréchet's results were generalized by Kryloff and Bogolioùboff (I) to a form which is substantially identical to Theorem 7.5 (ii) (cf. the note above on the hypotheses of the present discussion) but less general than Theorem 6.5 (ii).

Proof of (i). Let $p(x)$ be an absolute probability density corresponding to the given conditional probability density. For each fixed value of y for which

* Cf. the note to Theorem 7.1 (i).

$p(y) < \infty$, $p(x_0, y)$ is a P-measurable function, depending only on x_0, which is integrable on Ω; namely

$$\int p(x_0, y)dP = \int p(x, y)p(x)dx = p(y).$$

Then according to Theorem 6.1 (if $\phi(\omega) = p(x_0, y)$) the limit in (7.40) exists almost everywhere on Ω.

Proof of (ii). This proof follows the outline of the special case considered in Theorem 7.1 (ii). There is an X-measurable function $\phi(x)$, such that

$$\phi(x) > 0, \qquad \int \phi(x)dx = 1.*$$

If in (7.39) we define $Q(E)$ as $\int_E \phi(x)dx$, an absolute probability density $p(x)$ is obtained in the form

$$(7.39') \qquad \int_E p(x)dx = \lim_{\nu\to\infty} \frac{1}{N_\nu} \sum_{m=1}^{N_\nu} \int \phi(x)dx \int_E p^{(m)}(x, y)dy,$$

where the sequence $\{N_\nu\}$ is independent of E. We shall suppose that $p(x)$ is defined by (7.39'). If we define $\Pi^{(N)}(x, y)$ by

$$\Pi^{(N)}(x, y) = \frac{1}{N} \sum_{m=1}^{N} p^{(m)}(x, y),$$

we find (cf. equation (7.6)) that

$$(7.43) \qquad \frac{\nu + \mu}{\mu} \Pi^{(\nu+\mu)}(x, y) - \frac{\nu}{\mu} \Pi^{(\nu)}(x, y) = \int p^{(\nu)}(x, z)\Pi^{(\mu)}(z, y)dz.$$

Let E_0 be the X-set on which $p(x) = 0$. Then if $E \in F_x$, the limit in (7.41) exists, according to Theorem 6.1 (ii), for x almost everywhere in the complement of E_0, and

$$(7.44) \qquad \lim_{N\to\infty} \int_E \Pi^{(N)}(x, y)dy = q(x, E), \qquad x \in CE_0,$$

except possibly for an X-set of measure 0. According to equation (7.41),

$$\lim_{\nu\to\infty} \int \phi(x)dx \int_{E_0} \Pi^{(N)}(x, y)dy = \lim_{\nu\to\infty} \frac{1}{N_\nu} \sum_{m=1}^{N_\nu} \int \phi(x)dx \int_{E_0} p^{(m)}(x, y)dy = 0.$$

* If $\int 1dx = \lambda < \infty$, we can take $\phi(x) \equiv 1/\lambda$. Otherwise we use the fact (cf. §5) that there is an increasing sequence $E_1 \subset E_2 \subset \cdots$ of X-measurable sets such that $\sum_1^\infty E_i = X$ and that $\int_{E_j} 1dx = \lambda_j < \infty$. There is a sequence of positive numbers $\{\epsilon_n\}$ such that $\epsilon_0 + \sum_1^\infty \epsilon_n(\lambda_{n+1} - \lambda_n) = 1$, and we define $\phi(x)$ as ϵ_0 on E_1 and ϵ_n on $E_{n+1} - E_n$, for $n > 1$.

This implies that

$$\liminf_{m\to\infty} \int \phi(x)dx \int p^{(m)}(x, y)dy = 0. \tag{7.45}$$

Equation (7.45) means that some subsequence $\{\phi(x)\int_{E_0}p^{(a_j)}(x, y)\}$ of $\{\phi(x)\int_{E_0}p^{(m)}(x, y)dy\}$, when integrated over X, converges to 0. This implies that $\phi(x)\int_{E_0}p^{(a_j)}(x, y)dy$ converges in measure to 0* which in turn implies that a further subsequence $\{\phi(x)\int p^{(b_j)}(x, y)dy\}$ converges to 0 for almost all x. Since $\phi(x)>0$,

$$\lim_{j\to\infty} \int_{E_0} p^{(b_j)}(x, y)dy = 0$$

for almost all x. Now

$$\int_{E_0} p^{(b_j+1)}(x, y)dy = \int p(x, z)dz \int_{E_0} p^{(b_j)}(z, y)dy.$$

The z-integrand (for x fixed), $p(x, z)\int p^{(b_j)}(z, y)dy$, converges to 0, for almost all z, according to what has been just shown, and is less than or equal to the z-integrable function $p(x, z)$. Then, by a well known integration theorem, we can go to the limit under the integral sign, so that

$$\lim_{j\to\infty} \int_{E_0} p^{(b_j+1)}(x, y)dy = 0, \qquad x \in X. \tag{7.46}$$

If both sides of (7.43) are integrated with respect to y over a set E in the field, then if $\mu\to\infty$, we obtain

$$\begin{aligned}\liminf_{N\to\infty} \int_E \Pi^{(N)}(x, y)dy &= \liminf_{\mu\to\infty} \int p^{(\nu)}(x, z)dz \int_E \Pi^{(\mu)}(z, y)dy \\ &\geqq \liminf_{\mu\to\infty} \int_{CE_0} p^{(\nu)}(x, z)dz \int_E \Pi^{(\mu)}(z, y)dy.\end{aligned} \tag{7.47}$$

Now the z-integrand $p^{(\nu)}(x, z)\int_E \Pi^{(\mu)}(z, y)dy$ converges, as $\mu\to\infty$, to $p^{(\nu)}(x, z)q(z, E)$ for almost all z in CE_0, according to (7.44). Moreover this integrand is less than or equal to the z-integrable function $p^{(\nu)}(x, z)$. Then we can go to the limit under the integral sign, and obtain

$$\liminf_{N\to\infty} \int_E \Pi^{(N)}(x, y)dy \geqq \int_{CE_0} p^{(\nu)}(x, z)q(z, E)dz, \qquad \nu = 1, 2, \cdots.$$

* Convergence in measure was defined and discussed by F. Riesz, Comptes Rendus de l'Académie des Sciences, Paris, vol. 148 (1909), pp. 1303–1305.

On the other hand,

$$\limsup_{N\to\infty}\int_E \Pi^{(N)}(x, y)dy = \limsup_{\mu\to\infty}\left\{\int_{E_0} p^{(\nu)}(x, z)dz\int_E \Pi^{(\mu)}(z, y)dy + \int_{CE_0} p^{(\nu)}(x, z)dz\int_E \Pi^{(\mu)}(z, y)dy\right\}$$

$$\leqq \int_{E_0} p^{(\nu)}(x, z)dz + \int_{CE_0} p^{(\nu)}(x, z)q(z, E)dz, \qquad \nu = 1, 2, \cdots. \tag{7.48}$$

Then

$$\limsup_{N\to\infty}\int_E \Pi^{(N)}(x, y)dy - \liminf_{N\to\infty}\int_E \Pi^{(N)}(x, y)dy \leqq \int_{E_0} p^{(\nu)}(x, z)dz, \qquad \nu = 1, 2, \cdots. \tag{7.49}$$

If ν is allowed to increase without limit through the sequence $\{b_i+1\}$ of (7.46), it follows that the limit in (7.41) exists for all x, as was to be proved.

Now in addition to the other hypotheses, suppose that for some integer μ, $p^{(\mu)}(x, y)$ is bounded in x for each y. The existence of the limit in (7.41) is readily seen to imply that if $f(x)$ is a bounded X-measurable function,

$$\lim_{N\to\infty}\frac{1}{N}\sum_{m=1}^{N}\int p^{(m)}(x, z)f(z)dz$$

exists for all x. If we take $f(z) = p^{(\mu)}(z, y)$, then

$$\lim_{N\to\infty}\frac{1}{N}\sum_{m=1}^{N}\int p^{(m)}(x, z)p^{(\mu)}(z, y)dz = \lim_{N\to\infty}\frac{1}{N}\sum_{m=\mu+1}^{N+\mu} p^{(m)}(x, y) = \lim_{N\to\infty}\frac{1}{N}\sum_{m=1}^{N} p^{(m)}(x, y)$$

exists for all x, y, as was to be proved.

A given function $p(x, y)$ may correspond to several temporally homogeneous processes. If all these processes are metrically transitive, the function $p(x, y)$ will be called metrically transitive. If none of these processes has angle variables, $p(x, y)$ will be said to have no angle variables. Otherwise $p(x, y)$ will be said to be not metrically transitive, or to have angle variables, as the case may be.

THEOREM 7.6. *The function $p(x, y)$ is metrically transitive if and only if*

(i) *there is only a single absolute probability density $p(x)$; or*

(ii) *$q(E, x)$ (or $q(x, y)$ if the hypotheses of the second part of Theorem* 7.5 (ii) *are satisfied) is independent of x; or*

(iii) *the integral equation*

$$\int \psi(x)p(x, y)dx = \psi(y), \tag{7.50}$$

in the X-measurable integrable function $\psi(x)$, has only a single linearly independent solution; or

(iv) *there are disjunct X-sets F_1, F_2 in F_x, of positive X-measure, such that*

$$p(x, y) = 0, \qquad x \in F_j, \qquad y \notin F_j, \qquad (j = 1, 2), \tag{7.51}$$

*if we neglect (x, y)-sets of (x, y)-measure 0.**

The various parts of this theorem are proved by exactly the methods of the proof of Theorem 7.2. In the case of metric transitivity, the limit $q(x, E)$ can be expressed simply by

$$q(x, E) = \int_E p(y)dy. \tag{7.52}$$

(This equation corresponds to the equation $q_{jk} = p_k$ for all j, k in case II, when the given matrix is metrically transitive.) As an example of the proofs used, we prove (iii).

Proof of (iii). If the function $p(x, y)$ is metrically transitive, let $p(x)$ be the uniquely determined probability density. Then (7.52) is true. If $\psi(x)$ is X-measurable and integrable, and satisfies (7.50), it follows that

$$\int \psi(x)dx \int_E \Pi^{(N)}(x, y)dy = \int_E \psi(y)dy, \qquad N = 1, 2, \cdots;$$

and if $N \to \infty$, this becomes†

$$\int \psi(x)dx \int_E p(y)dy = \int_E \psi(y)dy. \tag{7.53}$$

If α is defined as $\int \psi(x)dx$, (7.53) implies, since E is arbitrary, that $\psi(y) = \alpha p(y)$ for almost all y. Since the functions $p(y)$, and $\psi(y)$ satisfy their integral equation identically, $\psi(y) \equiv \alpha p(y)$, as was to be proved. Conversely, if there is a solution of (7.52), uniquely determined (up to a constant factor), then the absolute probability density, which is a solution, is uniquely determined; hence there is metric transitivity, according to part (i).

* The equivalence of (i), (ii), (iii) was proved by Fréchet (I), in the case (described above) which he considered. The equivalence of (ii) and (iv) was announced by Kryloff and Bogolioùboff (II), whose hypotheses apparently exclude the possibility of exceptional values in (7.51).

† The x-integrand $\psi(x)\int_E \Pi^{(N)}(x, y)dy$ converges for all x to $\psi(x)\int_E p(y)dy$ and is uniformly less than or equal to $\psi(x)$; so we can go to the limit under the x-integral sign.

There is some interest in developing this theorem further.* Suppose a function $p(x, y)$ is not metrically transitive. Then X sets F_1, F_2 exist as described in (iv). Now it may be that F_i itself contains X-measurable sets F_{i1}, F_{i2}, of positive measure, such that $p(x, y) = 0$ if $x \in F_{ij}$, $y \notin F_{ij}$, if we neglect (x, y)-sets of (x, y)-measure 0. In the contrary case the function $p(x, y)$, considered only for $x \in F_i$, $y \in F_i$, is metrically transitive. Now it is readily seen that the uniform integrability condition prevents the existence of infinitely many X-measurable sets $F_1, F_2, \cdots$, of positive X-measure, such that if $x \in F_i$, $y \notin F_i$, then $p(x, y) = 0$ neglecting (x, y)-sets of (x, y)-measure 0. Hence there is at most a finite number μ of such sets, and $p(x, y)$ is metrically transitive when considered defined only for $x, y \in F_i$. Let $p_i(x)$ be the corresponding uniquely defined absolute probability density, and define $p_i(x) = 0$ if $x \notin F_i$. Then if $\rho_1, \cdots, \rho_\mu$ are non-negative numbers with sum 1, $p(x) = \sum_{i=1}^{\mu} \rho_i p_i(x)$ is an absolute probability density for $p(x, y)$, and conversely, any absolute probability density for $p(x, y)$ is such a linear combination. The limit $q(x, E)$ of Theorem 7.6 (ii) must be $\sum_{i=1}^{n} \rho_i \int_{E \cdot E_i} p_i(x)dx$.

THEOREM 7.7. *The function $p(x, y)$ has no angle variables if and only if*

(i) *whenever $E \in F_x$,*

$$\lim_{m \to \infty} \int_E p^{(m)}(x, y)dy = q(x, E) \tag{7.54}$$

exists for all $x \in X$ (or, in case the hypotheses of the second part of Theorem 7.5 (ii) are satisfied, whenever $\lim_{m \to \infty} p^{(m)}(x, y)$ exists for all x, y); or

(ii) *it is impossible to find disjunct-sets $E_1, \cdots, E_\nu$, $\nu > 1$, of positive X-measure, such that (if we neglect an (x, y)-set of (x, y)-measure 0),*

$$p(x, y) = 0, \quad \begin{cases} x \in E_r, & y \notin E_{r+1}, \quad r = 1, \cdots, \nu - 1, \\ x \in E_\nu, & y \notin E_1. \end{cases} \tag{7.55}$$

In case II, there seems to be no essential difference between the existence of angle variables and the existence of solutions (not equal to 1, of modulus 1) of the characteristic equation of the given matrix. However, in the present case it seems possible to obtain more general results by considering angle variables rather than solutions of the integral equation

$$\int \psi(y) p(x, y)dy = c\psi(x).$$

The greater adaptability of angle variables is shown, for example, by the fact

* Cf. Kryloff und Bogolioùboff (I, II), Doeblin and Fortet (I).

that the existence of an angle variable implies the existence of a bounded angle variable (as we have seen above). Fréchet was able to extend the usual Fredholm theory of integral equations to his kernels $p(x, y)$, and so could obtain the complete analogue of Theorem 7.3; and in the present treatment also, if the Fredholm theory is available, the proof of Theorem 7.3 goes right through in case III.

Proof of (i). Suppose that $p(x, y)$ has no angle variables. Let $p(x)$ be an absolute probability density corresponding to $p(x, y)$. We show first that the hypotheses of Theorem 6.5 are satisfied, so that there is a function $\phi(x, y)$ such that for every set $E \in F_x$ and for all x (except perhaps values in a set on which the integral of $p(x)$ vanishes),

$$\int_E p(x, y)dy = \int_E \phi(x, y)p(y)dy. \tag{7.56}$$

Let E_0 be the x-set on which $p(x)=0$. Since

$$\int p(x)dx \int_{E_0} p(x, y)dy = \int_{E_0} p(y)dy = 0, \tag{7.57}$$

$\int_{E_0} p(x, y)dy=0$, except possibly on an x-set on which the integral of $p(x)$ vanishes. Then if $\phi(x, y)$ is defined by

$$\begin{aligned} &\phi(x, y) = p(x, y)/p(y), && p(y) > 0 \\ &\phi(x, y) = 0, && p(y) = 0, \end{aligned} \tag{7.58}$$

it is readily verified that (7.56) holds, except possibly for values of x for which the integral of $p(x)$ vanishes. The hypotheses of Theorem 6.5 are therefore satisfied, and as the process has, by hypothesis, no angle variables, whatever absolute probability density is chosen, the limit in (7.56) must exist almost everywhere on CE_0. A suitable generalization of the proof of the corresponding part of Theorem 7.3 then completes the proof. Conversely, if the limit in (7.54) exists for all x, Theorem 6.5 states that the function $p(x, y)$ has no angle variables. The transition from (7.54) to the unintegrated form $(\lim_{m\to\infty} p^{(m)}(x, y))$ is easily made as in the proof of Theorem 7.5

Proof of (ii). If there are sets $E_1, \cdots, E_r$ as described in (ii), an angle variable can easily be explicitly defined, or the proof of the corresponding part of Theorem 7.3 can be generalized to show that $p(x, y)$ must have angle variables. Conversely suppose there is an angle variable, so that (cf. Theorem 6.3) there is an X-measurable function $\psi(x)$ such that

$$\psi(x_1) = c\psi(x_0), \qquad |c| = 1, c \neq 1, \tag{7.59}$$

if we neglect an (x_0, x_1)-set of P-measure 0. Then

$$\int [\psi(x_1) - c\psi(x_0)]P(de_{01}) = \int\int p(x)p(x, y)[\psi(y) - c\psi(x)]dx\,dy = 0,$$

so that

$$p(x)p(x, y)[\psi(y) - c\psi(x)] = 0 \tag{7.60}$$

for almost all (x, y), and

$$p(x, y)[\psi(y) - c\psi(x)] = 0, \qquad x \notin E_0, \tag{7.61}$$

if we neglect (x, y)-sets of zero measure. Let ξ be a point not in E_0, such that

$$\int_{E_0} p(\xi, y)dy = p(\xi, y)[\psi(y) - c\psi(\xi)] = 0$$

(where the second is to hold for almost all values of y). This may exclude (cf. (7.5)) a ξ-set of measure 0, besides E_0. Then, since $p(\xi, y) > 0$ on CE_0 on a set of positive y-measure (its integral over CE_0 is 1), $\psi(x_1)$ takes on the value $c\psi(\xi)$ on a set of positive P-measure. Now let a be any value, not equal 0, assumed by ψ on a set of positive P-measure. According to Theorem 6.3 (iii), c must be a root of unity. Let ν be the smallest exponent r for which $c^r = 1$. The function $\psi(x)$ takes on values $a, ca, \cdots, c^{\nu-1}a$ on subsets $E_1, \cdots, E_\nu$ of CE_0. The fact that if $x_0 \in E_r$, then $x_1 \in E_{r+1}$, $(r = 1, \cdots, \nu-1)$, (if $x_0 \in E_\nu$, then $x_1 \in E_1$) necessarily, if we neglect sets of P-measure 0, and that the set determined by the condition $x_0 \in E + \cdots + E_\nu$ is invariant up to a set of P-measure 0, implies the conditions of (7.55).

The set $E_1 + \cdots + E_\nu$ is one of the sets F (corresponding to invariant Ω-sets) analyzed above. In general there will be then a finite number of F-sets, and if the function $p(x, y)$ has any angle variables, one or more of these F-sets will be divided into a finite number of E-sets.*

Combining the previous theorems we obtain finally the theorem:

THEOREM 7.8. *The function $p(x, y)$ is metrically transitive and has no angle variables if and only if*

(i) *whenever $E \in F_x$, $\lim_{m\to\infty} \int_E p^{(m)}(x, y)dy$ exists for all x and is independent of x (or, in case the conditions of the second part of Theorem* 7.5 (ii) *are satisfied, if $\lim_{m\to\infty} p^{(m)}(x, y)$ exists for all x, y and is independent of x); or*

(ii) *it is impossible to find sets as described in Theorems* 7.6 (iv) *or* 7.7 (ii); *or*

(iii) *every function $p(x, y)$, $p^{(2)}(x, y)$, $\cdots$ is metrically transitive.*

* This decomposition of X was announced by Doeblin and Fortet (I). The hypothesis, made here, that absolute and conditional probabilities are given by density functions, is unnecessary, as we only need enough hypotheses to assure the fact that an angle variable will assume one of its values on a set of positive P-measure.

IV. The case requires little comment. The properties of the transformation S correspond to similar properties of T; for example, if one has angle variables, so has the other.

BIBLIOGRAPHY

N. BOGOLIOÙBOFF.
Cf. N. Kryloff and N. Bogolioùboff below.

P. J. DANIELL.
(I) *Integrals in an infinite number of dimensions*, Annals of Mathematics, (2), vol. 20 (1918–1919), pp. 281–288.

W. DOEBLIN and R. FORTET.
(I) *Sur deux notes de Mm. Kryloff et Bogolioùboff*, Comptes Rendus de l'Académie des Sciences, Paris, vol. 204 (1937), pp. 1699–1701.

J. L. DOOB.
(I) *Probability and statistics*, these Transactions, vol. 36 (1934), pp. 759–775.
(II) *Stochastic processes depending on a continuous parameter*, these Transactions, vol. 42 (1937), pp. 107–140.

R. FORTET.
Cf. W. Doeblin and R. Fortet.

M. FRÉCHET
(I) *Compléments à la théorie des probabilités discontinues "en chaîne,"* Annali delle Università Toscane, Pisa, (2), vol. 2 (1933), pp. 131–164.
(II) *Sur l'allure asymptotique des densités itérées dans le problème des probabilités "en chaîne,"* Bulletin de la Société Mathématique de France, vol. 62 (1934), pp. 68–83.

M. FRÉCHET and J. HADAMARD
(I) Zeitschrift für angewandte Mathematik und Mechanik, vol. 13 (1933), pp. 92–97.

J. HADAMARD.
Cf. M. Fréchet and J. Hadamard.

E. HOPF.
(I) *On causality, statistics and probability*, Journal of Mathematics and Physics of the Massachusetts Institute of Technology, vol. 13 (1934), pp. 51-102.

B. HOSTINSKY.
(I) *Application du calcul des probabilités à la théorie du mouvement Brownien*, Annales de l'Institute Henri Poincaré, vol. 3 (1932), pp. 1–74.
(II) *Méthodes générales du calcul des probabilités*, Mémorial des Sciences Mathématiques, no. 52.

M. KAC.
(I) *Sur les fonctions indépendantes* I, Studia Mathematica, vol. 6 (1936), pp. 46–58.

A. KHINTCHINE.
(I) *Zur mathematischen Begründung der statistischen Mechanik*, Zeitschrift für angewandte Mathematik und Mechanik, vol. 13 (1933), pp. 101–103.

A. KOLMOGOROFF.
(I) *Über die analytische Methoden in der Wahrscheinlichkeitsrechnung*, Mathematische Annalen, vol. 104 (1931), pp. 415–458.
(II) *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Ergebnisse der Mathematik, vol. 2, no. 3.

N. KRYLOFF and N. BOGOLIOÙBOFF.
(I) *Sur les probabilités en chaîne*, Comptes Rendus de l'Académie des Sciences, Paris, vol. 204 (1933), pp. 1386–1388.
(II) *Les propriétés ergodiques des suites des probabilités en chaîne*, Comptes Rendus de l'Académie des Sciences, Paris, vol. 204 (1937), pp. 1454–1456.

P. LÉVY.
(I) *Théorie de l'Addition des Variables Aléatoires*, Paris, 1937.

Z. ŁOMNICKI and S. ULAM.

(I) *Sur la théorie de la mesure dans les espaces combinatoires et son application au calcul des probabilités*, I *Variables indépendantes*, Fundamenta Mathematicae, vol. 23 (1934), pp. 237–278.

R. VON MISES.

(I) *Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik*, Leipzig and Vienna, 1931.

R. E. A. C. PALEY and N. WIENER.

(I) *Fourier Transforms in the Complex Domain*, American Mathematical Society Colloquium Publications, vol. 19, 1934.

V. ROMANOVSKY.

(I) *Recherches sur les chaînes de Markoff*, Acta Mathematica, vol. 66 (1936), pp. 147–251.

S. ULAM.

Cf. Z. Łomnicki and S. Ulam.

N. WIENER.

Cf. R. E. A. C. Paley and N. Wiener.

UNIVERSITY OF ILLINOIS,
URBANA, ILL.

CORRECTION TO THE PAPER "THE MULTINOMIAL SOLID AND THE CHI TEST"*

BY
BURTON H. CAMP

The following note should be appended at the bottom of page 141:†

However, Pearson's sum (loc. cit.) extends over the interior of an ellipsoid in $(m-1)$-space, instead of over the interior of our parallelopiped. The two figures overlap, but do not coincide; hence it should not be asserted, as in §5, that the estimated error surely applies to the Pearson tables. An attempt to extend the theory further has encountered so many difficulties that the author would be compelled at present to adopt the error found in §5 as his best estimate of the error inherent also in Pearson's approximation. But it might not always be a close estimate. The equations (20) and the restrictions (7) do apply to both cases. If an expression were to be found for the sum Q, when extended over the interior of our parallelopiped, one would have a test of significance which would rival Pearson's chi-square test, and which, unlike his, would reduce to the ordinary procedure when $m=2$ (point binomial). It would give the probability that a sample $(f_1, \cdots, f_m)$ had class frequencies as near the ideal as those observed, whereas Pearson's test gives the probability of a sample's having class frequencies whose total probability would be as great as the total probability of the sample observed. These two possible tests are not identical. The form just suggested is the natural extension of the method usually preferred for the point binomial, and the estimated error could be made to apply to it with exactness; but this estimate is still useful as an approximation in Pearson's case, and the restrictions (7) continue to be needed.

* Received by the editors December 20, 1937.

† These Transactions, vol. 31 (1929), pp. 133–144.

WESLEYAN UNIVERSITY,
MIDDLETOWN, CONN.

GEOMETRY IN AN n-DIMENSIONAL SPACE WITH THE ARC LENGTH

$$s = \int \{A_i(x, x')x''^i + B(x, x')\}^{1/p}dt^*$$

BY

AKITSUGU KAWAGUCHI

Geometry in the manifold $K_n^{(m)}$ of line elements of a higher order, that is, $(x, dx, \cdots, d^m x)$ or $(x, x', \cdots, x^{(m)})$, in a space of n dimensions with a point coordinate system x^i, $(i=1, 2, \cdots, n)$, was first furnished with its foundation by the present author.† In close connection with this theory, there are two important problems. The first is the geometry of paths in this space, that is, the geometrical study of the system of paths $x^i=x^i(t)$ which are defined by the ordinary differential equations of the form

$$x^{(m+1)i} + \Gamma^i(t, x, x', \cdots, x^{(m)}) = 0,$$

where $x^{(\lambda)i}$ denotes $d^\lambda x^i/dt^\lambda$ and t is a parameter. The second is a case of metric geometry in this space, that is, the study of the geometrical properties characterized by a metric which gives as the arc length s of a curve $x^i=x^i(t)$

$$s = \int F(t, x, x', \cdots, x^{(m)})dt.$$

The first problem is a generalization of the so-called geometry of paths in a space with a generalized affine connection‡ $x''^i+\Gamma^i(t, x, x')=0$, and was

* Presented to the Society, December 28, 1937; received by the editors February 8, 1937, and, in revised form, August 3, 1937.

A summary of the present paper has already appeared in Proceedings of the Imperial Academy, vol. 12 (1936), pp. 205–208.

† A. Kawaguchi, *Die Differentialgeometrie in der verallgemeinerten Mannigfaltigkeit*, Rendiconti del Circolo Matematico di Palermo, vol. 56 (1932), pp. 245–276; *Theory of connections in the generalized Finsler manifold*, Proceedings of the Imperial Academy, vol. 7 (1931), pp. 211–214; ibid., vol. 8 (1932), pp. 340–343; ibid., vol. 9 (1933), pp. 347–350.

‡ L. Berwald, *Untersuchung über die Krümmung allgemeiner metrischer Räume auf Grund des in ihnen herrschenden Parallelismus*, Mathematische Zeitschrift, vol. 25 (1926), pp. 40–73; J. Douglas, *The general geometry of paths*, Annals of Mathematics, (2), vol. 29 (1927), pp. 143–168; D. D. Kosambi, *Parallelism and path-space*, Mathematische Zeitschrift, vol. 37 (1933), pp. 608–618, and E. Cartan, *Observations sur le memoire précédent*, ibid., pp. 619–622; D. D. Kosambi, *Systems of differential equations of the second order*, Quarterly Journal of Mathematics (Oxford), vol. 6 (1935), pp. 1–12; W. Ślebodziński, *Sur deux connections affines généralisées*, Prace Matematyczno-Fizyczne, vol. 43 (1935), pp. 1–39.

BOSTON UNIVERSITY
COLLEGE OF LIBERAL ARTS
LIBRARY

first propounded by D. D. Kosambi.* Although he derived many interesting results, his theory seems to be systematically not yet complete, and there are many irregularities. The present writer and H. Hombu have reconstructed this theory from another and more general standpoint, by which the irregularities of the theory of Kosambi exhibit themselves.†

On the other hand, the second problem is a generalization of a Finsler space and there have been many attempts to solve it,‡ but it remains yet unsolved. Recently E. Cartan§ has introduced a connection, which is related intrinsically with an integral $\int F(x, y, y', y'')dx$, in a plane under the group of all contact transformations by use of several invariant Pfaffians and has discussed in detail a special case where the integral has the form $\int(Ay''+B)^{p}dx$. This theory of Cartan has been extended by H. Hombu‖ to the case where the integral possesses the form $\int F(x, y, y', y'', y''')dx$. These theories give a foundation to the invariant theory of the integral under contact transformations and take the first step towards the general problem. But unfortunately, by reason of its restricted method which is available only for a plane, these theories cannot be extended to an n-dimensional space as they stand. Furthermore it is more desirable to set up the theory under the group of all point transformations rather than contact transformations.

In the present work it is proposed to establish the foundation of the geometry in an n-dimensional space with arc length given by the integral $\int\{A_i(x, x')x''^{i}+B(x, x')\}^{1/p}dt$, by introducing several connections under the group of all point transformations. Although the integral has a special form, this would be the second step towards the general problem; and in this special case many concrete results are obtainable, which may be of interest. The integrand does not contain the parameter t explicitly, and I shall investigate mainly only those properties that are invariant under any (analytic) transformation of the parameter t, for the reason that only these properties have geometrical meaning, that is, are related intrinsically to the base curves. This

* D. D. Kosambi, *Path-spaces of higher order*, Quarterly Journal of Mathematics (Oxford), vol. 7 (1936), pp. 97–104.

† A. Kawaguchi and H. Hombu, *Die Geometrie des Systems der partiellen Differentialgleichungen höherer Ordnung*, Journal of the Faculty of Science, Hokkaido Imperial University, vol. 6 (1937), pp. 21–62.

‡ H. V. Craig, *On a generalized tangent vector*, American Journal of Mathematics, vol. 57 (1935), pp. 457–462; J. L. Synge, *Some intrinsic and derived vectors in a Kawaguchi space*, ibid., pp. 679–691; A. Kawaguchi, *Some intrinsic derivations in a generalized space*, Proceedings of the Imperial Academy, vol. 12 (1936), pp. 149–151; *Certain identities in a generalized space*, ibid., pp. 152–155.

§ E. Cartan, *La géométrie de l'intégrale* $\int F(x, y, y', y'')dx$, Journal de Mathématiques, (9), vol. 15 (1936), pp. 42–69.

‖ H. Hombu, *Invariantentheorie des Integrals* $\int F(x, y, y', y'', y''')dx$, Proceedings of the Imperial Academy, vol. 12 (1936), pp. 156–158; *Geometrie des Integrals* $\int(Ly'''+M)dx$, ibid., pp. 159–161.

restriction makes the theory somewhat complicated. The theory can be generalized without difficulty to the case where the integral is of the form

$$\int \{A_i(x, x', x'')x'''^i + B(x, x', x'')\}^{1/p}dt^*$$

or

$$\int \{a_i(x, x')a_j(x, x')x''^i x''^j + 2b(x, x')a_i x''^i + c(x, x')\}^{1/p}dt.\dagger$$

Chapter 1 is devoted to discussion of the general case, where $p \neq 3/2$. In this case there exists a system of paths having the equations of the form $x''^i + \Gamma^i(x, x') = 0$, which leads to the definition not only of the base connection but also of a connection C, by recalling the geometry of paths in the manifold $K_n^{(1)}$. There is another connection C' which is stated in Chapter 2 and which has a form similar to that of Cartan in a Finsler space.‡

Chapter 1. Base connection and connection C

1. Let us consider an n-dimensional metric space such that the arc length of a curve $x^i = x^i(t)$ is given by the integral

$$s = \int \{A_i(x, x')x''^i + B(x, x')\}^{1/p}dt, \tag{1.1}$$

where $x'^i = dx^i/dt$, $x''^i = d^2x^i/dt^2$, and A_i and B are differentiable functions of x^i and x'^i. In order that the arc length should be related intrinsically to the curve, that is, that it should remain unaltered by a transformation of the parameter t, we must have

$$A_i x'^i = 0, \tag{1.2}$$

$$2A_i x''^i + (A_{k(i)}x''^k + B_{(i)})x'^i = p(A_i x''^i + B), \tag{1.3}$$

where

$$A_{k(i)} = \frac{\partial A_k}{\partial x'^i}, \qquad B_{(i)} = \frac{\partial B}{\partial x'^i}.$$

We shall denote partial differentiation by x'^i with (i) and that by x^i with i throughout this paper. Equation (1.3) implies

---

* T. Ohkubo, *Base connections in a special Kawaguchi space*, Journal of the Faculty of Science, Hokkaido Imperial University, vol. 5 (1936), pp. 167–188.

† H. Hokari, *Die Geometrie des Integrals* $\int(a_i a_j x''^i x''^j + 2ba_i x''^i + c)^{1/p}dt$, Proceedings of the Imperial Academy, vol. 12 (1936), pp. 209–212.

‡ E. Cartan, *Les espaces de Finsler*, Actualités Scientifiques et Industrielles, vol. 79, 1934.

$$A_{k(i)}x'^i = (p-2)A_k, \qquad B_{(i)}x'^i = pB. \tag{1.4}$$

Thus, the A_i are homogeneous of degree $p-2$ with regard to the x'^i, and B is homogeneous of degree p. From (1.2) it follows, on differentiating by x'^j, that

$$A_{i(j)}x'^i = -A_j. \tag{1.5}$$

2. The point transformation $y^i = y^i(x^j)$ gives rise to the following relations:

$$y'^r = P_s^r x'^s, \qquad y''^r = P_s^r x''^s + P_{sq}^r x'^s x'^q, \tag{1.6}$$

where

$$P_s^r = \frac{\partial y^r}{\partial x^s}, \qquad Q_s^r = \frac{\partial x^r}{\partial y^s} \qquad \text{and} \qquad P_{sq}^r = \frac{\partial^2 y^r}{\partial x^s \partial x^q}.$$

Since the arc length (1.1) is a scalar, it can be seen at once that A_i is a vector. But B is not a scalar and becomes, by the point transformation,

$$B(y, y') = B(x, x') - A_q(x, x')Q_i^q P_{jk}^i x'^j x'^k. \tag{1.7}$$

Besides A_i we get two other intrinsic vectors

$$T_i = (A_{k(i)} - 2A_{i(k)})x''^k - 2A_{il}x'^l + B_{(i)}, \tag{1.8}$$

$$\begin{aligned} E_i = {} & (A_{i(k)} - A_{k(i)})x'''^k + (A_{i(k)(l)} - A_{k(i)(l)})x'''^l x''^k \\ & + (A_{i(k)l}x'^l - A_{k(i)l}x'^l + A_{ki} + A_{ik} - B_{(i)(k)} + A_{ij(k)}x'^j)x''^k \\ & + A_{ijk}x'^j x'^k - B_{(i)l}x'^l + B_i. \end{aligned} \tag{1.9}$$

The vector T_i is the Craig vector*

$$T_i = -2\frac{d}{dt}\frac{\partial F}{\partial x''^i} + \frac{\partial F}{\partial x'^i},$$

and E_i is the Euler vector

$$E_i = \frac{d^2}{dt^2}\frac{\partial F}{\partial x''^i} - \frac{d}{dt}\frac{\partial F}{\partial x'^i} + \frac{\partial F}{\partial x^i}$$

of the function $F = A_i x''^i + B$. It can be verified easily that

$$T_i x'^i = p(A_i x''^i + B), \qquad E_i x'^i = (1-p)\frac{d}{dt}(A_i x''^i + B). \tag{1.10}$$

---

* See H. V. Craig, *On a generalized tangent vector*, loc. cit.

When s stands for t, we get the normalized vectors T_i^* and E_i^* for T_i and E_i. Then we have

$$T_i^*\dot{x}^i = p, \qquad E_i^*\dot{x}^i = 0, \tag{1.10'}$$

where the dot denotes differentiation by s. Among these vectors there are the relations

$$\begin{aligned} T_i^* &= T_i\left(\frac{dt}{ds}\right)^{p-1} - (2p-3)A_i\left(\frac{dt}{ds}\right)^{p-3}\frac{d^2t}{ds^2}, \\ E_i^* &= E_i\left(\frac{dt}{ds}\right)^{p} - (p-1)T_i\left(\frac{dt}{ds}\right)^{p-2}\frac{d^2t}{ds^2} \\ &\quad + (p-1)A_i\left\{(p-3)\left(\frac{d^2t}{ds^2}\right)^2 + \frac{dt}{ds}\frac{d^3t}{ds^3}\right\}\left(\frac{dt}{ds}\right)^{p-4}. \end{aligned} \tag{1.11}$$

3. If $2p=3$, the determinant of the tensor

$$G_{ik} = 2A_{i(k)} - A_{k(i)} \tag{1.12}$$

vanishes, identically, because

$$G_{ik}x'^k = 0.$$

Now we assume that $p\neq 3/2$, and that the determinant does not vanish identically; then we have from (1.8)

$$x^{[2]l} \equiv -T_iG^{il} = x''^l + 2\Gamma^l, \tag{1.13}$$

where

$$2\Gamma^l = (2A_{ik}x'^k - B_{(i)})G^{il} \qquad \text{and} \qquad G_{ik}G^{il} = \delta_k^l. \tag{1.14}$$

The functions Γ^l are homogeneous of degree two with regard to the x'^i. As the relation $G_{ik}x'^k=(2p-3)A_i$ holds, the last equation of (1.14) shows that

$$A_iG^{il} = \frac{1}{2p-3}x'^l. \tag{1.15}$$

From $G_{ki}G^{kl}=\delta_i^l$ we get also

$$A_iG^{il} = -\frac{1}{p}x'^i, \tag{1.15'}$$

which leads to

$$F = A_ix^{[2]i}. \tag{1.16}$$

From (1.14) we obtain further

$$2\Gamma^{i}A_{i} = B. \tag{1.17}$$

Now we shall introduce, as in a Finsler space, the manifold $K_{n}^{(1)}$ of line elements (x, x') in the space; and for this manifold (1.13) gives the base connection in our space.*

4. Let v^{i} be a contravariant vector field in $K_{n}^{(1)}$, homogeneous of degree zero with regard to the x'^{i}, that is, independent of a choice of the parameter t. Then a new vector can be derived from v^{i} in connection with the base connection:

$$\frac{\delta v^{i}}{dt} = \frac{dv^{i}}{dt} + \Gamma^{i}_{(j)}v^{j}, \dagger \tag{1.18}$$

which is an absolute derivative of the vector v^{i} along the curve and is independent of transformation of t. From (1.18) we can define further the absolute differential corresponding to a displacement from a line element (x, x') to a neighboring line element $(x+dx, x'+dx')$

$$\delta v^{i} = dv^{i} + \Gamma^{i}_{(j)(k)}v^{j}dx^{k}. \tag{1.19}$$

This absolute differential remains unaltered by any change of t. By any transformation of coordinates $\Gamma^{i}_{(j)(k)}$ are transformed in the same way as the parameters of an affine connection and are homogeneous of degree zero with regard to the x'^{h}, that is, they are invariant under transformation of t. For this reason the connection C, defined by (1.19), characterizes the geometrical properties of our space. It is not necessary, therefore, to take the line elements (x, x') along any curve; they can be taken arbitrarily at any point.

On account of the base connection (1.13) we have for a displacement of a line element

$$\delta x'^{i} = dx'^{i} + \Gamma^{i}_{(j)}dx^{j}. \tag{1.20}$$

On account of this, we can write (1.19) in the form

$$\begin{aligned}\delta v^{i} &= \frac{\partial v^{i}}{\partial x^{j}}dx^{j} + \frac{\partial v^{i}}{\partial x'^{j}}dx'^{j} + \Gamma^{i}_{(k)(j)}v^{k}dx^{j} \\ &= dx^{j}\nabla_{j}v^{i} + \delta x'^{j}\nabla'_{j}v^{i},\end{aligned} \tag{1.21}$$

where

* For the base connection reference may be made to A. Kawaguchi, *Die Differentialgeometrie in der verallgemeinerten Mannigfaltigkeit*, loc. cit.

† As $x^{[h]i}$ is a vector, we know from (1.13) that Γ^{i} behave under a coordinate transformation in the same way as affine connection parameters γ^{i}_{jk} multiplied by $x'^{j}x'^{k}$. From this fact it can be concluded that $\delta v^{i}/dt$ is a contravariant vector.

$$(1.22)\qquad \nabla_j v^i = \frac{\partial v^i}{\partial x^j} - \frac{\partial v^i}{\partial x'^k}\Gamma^k_{(j)} + \Gamma^i_{(k)(j)}v^k, \qquad \nabla'_j v^i = \frac{\partial v^i}{\partial x'^j},$$

which are defined as the covariant derivatives of a vector v^i. $\nabla'_j v^i$ are multiplied by a factor $dt/d\bar{t}$ for a change of the parameter $\bar{t}=\bar{t}(t)$. For a covariant vector or a tensor we can define covariant derivatives in the same manner as in usual tensor calculus, for example,

$$(1.23)\qquad \delta v_i = dv_i - \Gamma^j_{(i)(k)}v_j dx^k.$$

In particular we obtain for A_i

$$(1.24)\qquad \nabla_j A_i = A_{ij} - A_{i(k)}\Gamma^k_{(j)} - \Gamma^k_{(i)(j)}A_k, \qquad \nabla'_j A_i = A_{i(j)},$$

for which the following relations hold:

$$(1.25)\qquad x'^j\nabla_j A_i = 0, \qquad x'^i\nabla_j A_i = 0, \qquad \frac{\delta A_i}{dt} = A_{i(j)}x^{[2]j}.$$

These can be verified without difficulty from (1.4), (1.5), and (1.14).

The covariant differential of a relative tensor $\mathfrak{v}^i$ of weight k is also given as follows:

$$(1.26)\qquad \delta\mathfrak{v}^i = d\mathfrak{v}^i + \Gamma^i_{(j)(k)}\mathfrak{v}^j dx^k - k\mathfrak{v}^i\Gamma_{(j)}dx^j,$$

where $\Gamma_{(j)} \equiv \Gamma^t_{(t)(j)}$. Hence

$$(1.27)\qquad \nabla_j\mathfrak{v}^i = \frac{\partial\mathfrak{v}^i}{\partial x^j} - \frac{\partial\mathfrak{v}^i}{\partial x'^h}\Gamma^h_{(j)} + \Gamma^i_{(h)(j)}\mathfrak{v}^h - k\mathfrak{v}^i\Gamma_{(j)},$$
$$\nabla'_j\mathfrak{v}^i = \frac{\partial\mathfrak{v}^i}{\partial x'^j}.$$

5. If the components of a vector v^i are homogeneous of degree h with regard to the x'^j, then they are components of a geometrical object in some sense. But the δv^i in the sense of (1.19) have no geometrical meaning. That is, the δv^i are not independent of the parameter t but defined when and only when a choice of a parameter t is held fixed. In fact, by transformation of the parameter t one obtains

$$\delta\bar{v}^i = \left(\frac{dt}{d\bar{t}}\right)^h \delta v^i + hv^i\left(\frac{dt}{d\bar{t}}\right)^{h-1} d\frac{dt}{d\bar{t}},$$

which carries no geometrical meaning. Although for such a vector the absolute differential having geometrical meaning cannot be defined, it is possible to define it when the values of x''^i in the considered line element (x, x') are

given by any reason,† as we shall now show. Nevertheless it must be noticed that the covariant derivatives (1.22) of such a vector are surely homogeneous with regard to the x'^{j}.

Since x''^{i} are known in the considered line element (x, x'), the values of F and $x^{[2]i}$ are also determined. For displacement from a line element (x, x') to a neighboring line element $(x+dx, x'+dx')$ it is known that $\delta A_i x^{[2]i}$ and $(\delta A_i/dt)\delta x'^{i}$ are both scalars. They are varied by transformation of t as follows:

$$(\delta A_i x^{[2]i})\bar{} = \sigma^p \delta A_i x^{[2]i} + (p-2)\sigma^{p-1} F d\sigma - \sigma^{p-2}\frac{d\sigma}{d\bar{t}} A_i \delta x'^{i},$$

$$\left(\frac{\delta A_i}{dt}\delta x'^{i}\right)\bar{} = \sigma^p \frac{\delta A_i}{dt}\delta x'^{i} + (p-2)\sigma^{p-2}\frac{d\sigma}{d\bar{t}} A_i \delta x'^{i} - \sigma^{p-1} F d\sigma,$$

where $\sigma = dt/d\bar{t}$. Accordingly one gets

$$(\lambda_i dx^i + \mu_i \delta x'^{i})\bar{} = \lambda_i dx^i + \mu_i \delta x'^{i} - \frac{d\sigma}{\sigma},$$

where

$$\lambda_i = \frac{1}{(p-1)(p-3)F} x^{[2]k}\nabla_i A_k,$$

$$\mu_i = \frac{1}{(p-1)(p-3)F} x^{[2]k}\{(p-2)A_{k(i)} + A_{i(k)}\},$$

since

$$\lambda_i dx^i + \mu_i \delta x'^{i} = \frac{1}{(p-1)(p-3)F}\left\{(p-2)x^{[2]i}\delta A_i + \frac{\delta A_i}{dt}\delta x'^{i}\right\},$$

when $p \neq 1, 3$. Then one obtains as the required absolute differential

$$\delta^* v^i = \delta v^i + h v^i(\lambda_j dx^j + \mu_j \delta x'^{j}), \tag{1.28}$$

accompanied by the covariant derivatives

$$\begin{aligned} \nabla_j^* v^i &= \nabla_j v^i + \frac{h}{(p-1)(p-3)F} v^i x^{[2]k}\nabla_j A_k, \\ \nabla_j'^* v^i &= \nabla_j' v^i + \frac{h}{(p-1)(p-3)F} v^i x^{[2]k}\{(p-2)A_{k(j)} + A_{j(k)}\}, \end{aligned} \tag{1.29}$$

since

$$x'^{i}\nabla_k A_i = 0, \qquad \{(p-2)A_{k(j)} + A_{j(k)}\}x'^{k} = 0. \tag{1.30}$$

† For example, if one considers an absolute differential adjoint to a curve touching the line element (x, x'), one can take x''^{i} along the curve as the associated one.

$F^{h/p}\delta v^i$ are independent of a change of the parameter t. For x'^i equation (1.29) yields

$$\frac{\delta^* x'^i}{dt} = x^{[2]i} + \frac{1}{(p-3)F} x'^i A_{k(j)} x^{[2]k} x^{[2]j}. \tag{1.31}$$

If v^i be defined in the manifold $K_n^{(2)}$ of line elements of the second order (x, x', x'') and go into $\sigma^h v^i$ by transformation

$$\bar{x}'^i = \sigma x'^i,$$

$$\bar{x}''^i = \sigma^2 x''^i + \sigma \frac{d\sigma}{dt} x'^i,$$

then it is very easy to derive δv^i, which goes into $\sigma^h \delta v^i$ for the same transformations:

$$\delta v^i = dv^i + \Gamma^i_{(j)(k)} v^j dx^k - \frac{h}{p} v^i d \log F. \tag{1.32}$$

6. Displace a line element (x, x') parallel to itself in the sense of the base connection on its direction. A curve is obtained, which satisfies the differential equations

$$x^{[2]i} = 0, \tag{1.33}$$

that is,

$$x''^i + 2\Gamma^i = 0.$$

These differential equations define a system of paths in our space, but these paths are not extremal curves of the integral $\int F^{1/p} dt$, unlike those in the Finsler space. They are minimal curves, for along them $F = 0$ holds. But it is to be noticed that minimal curves do not always satisfy (1.33).

The equations in (1.33) in any parameter t become

$$x''^i + 2\Gamma^i = \rho x'^i,$$

where ρ is a function of t. We shall refer to a parameter t as *affine length*, if it makes ρ equal zero. It is determined except for a linear transformation $\bar{t} = at + b$. If we confine ourselves to affine length, the discussion in §6 is superfluous.

Considering the covariant derivatives of a vector v^i along any path, we obtain from (1.22)

$$\nabla_0 v^i = \frac{\partial v^i}{\partial x^j} x'^j - 2 \frac{\partial v^i}{\partial x'^j} \Gamma^j + \Gamma^i_{(j)} v^j, \tag{1.34}$$

as $\delta x'^i = 0$. Particularly it is known from (1.25) that

$$\nabla_0 A_i = 0. \tag{1.35}$$

The paths are formally the same ones as those in the space with a generalized affine connection. There is, however, a difference in that in our space there is an intrinsic covariant vector A_i besides the functions Γ^i, and our space is characterized by both quantities A_i and Γ^i. We can develop the theory of curvatures, treat the equivalence problem, and so forth, by slight modifications of the methods employed in the geometry of generalized paths.

7. Computing the parenthesis of Poisson for the covariant derivatives, we find the curvature tensors as follows:

$$\begin{aligned}(\nabla_j\nabla_k - \nabla_k\nabla_j)v^i &= -R^{\cdots i}_{jkl}v^l + K^{\cdots l}_{jk}\nabla'_l v^i,\\ (\nabla_j\nabla'_k - \nabla'_k\nabla_j)v^i &= -B^{\cdots i}_{jkl}v^l,\end{aligned} \tag{1.36}$$

where we put

$$\begin{aligned}B^{\cdots i}_{jkl} &= \Gamma^i_{(j)(k)(l)},\\ R^{\cdots i}_{jkl} &= \frac{\partial\Gamma^i_{(l)(j)}}{\partial x^k} - \frac{\partial\Gamma^i_{(l)(k)}}{\partial x^j} + \Gamma^h_{(l)(j)}\Gamma^i_{(k)(h)} - \Gamma^h_{(l)(k)}\Gamma^i_{(j)(h)}\\ &\quad + \Gamma^h_{(j)}\Gamma^i_{(l)(k)(h)} - \Gamma^h_{(k)}\Gamma^i_{(l)(j)(h)},\\ K^{\cdots i}_{jk} &= \frac{\partial\Gamma^i_{(j)}}{\partial x^k} - \frac{\partial\Gamma^i_{(k)}}{\partial x^j} + \Gamma^h_{(j)}\Gamma^i_{(k)(h)} - \Gamma^h_{(k)}\Gamma^i_{(j)(h)}.\end{aligned} \tag{1.37}$$

It is not difficult to find the relations satisfied by these curvature tensors. They are

$$\begin{aligned}&R^{\cdots i}_{jkl} + R^{\cdots i}_{kjl} = 0, \qquad R^{\cdots i}_{jkl} + R^{\cdots i}_{klj} + R^{\cdots i}_{ljk} = 0,\\ &K^{\cdots i}_{jk} = R^{\cdots i}_{jkl}x'^l, \qquad B^{\cdots i}_{jkl}x'^l = 0,\\ &R^{\cdots i}_{jkl} = K^{\cdots i}_{jk\cdot(l)} = \nabla'_l K^{\cdots i}_{jk}.\end{aligned} \tag{1.38}$$

The identities of Bianchi are now expressed by

$$\begin{aligned}\nabla_{[h}R^{\cdots i}_{jk]l} + K^{\cdots r}_{[hj}B^{\cdots i}_{k]lr} &= 0,\\ 2\nabla_{[h}B^{\cdots i}_{j]kl} + \nabla'_l R^{\cdots i}_{hjk} &= 0,\\ \nabla_{[h}K^{\cdots i}_{jk]} &= 0.\end{aligned} \tag{1.39}$$

The invariant theory under transformation of the parameter t can also be developed by the fundamental descriptive invariants of Douglas*

* J. Douglas, loc. cit.

$$\Pi^{i}_{.jk} = \Gamma^{i}_{(j)(k)} - \frac{\delta_j{}^i}{n+1}\Gamma^{h}_{(h)(k)} - \frac{\delta_k{}^i}{n+1}\Gamma^{h}_{(j)(h)} - \frac{1}{n+1}x'^i B^{\cdots h}_{jkh},$$

$$\mathfrak{H}^{\cdots i}_{jkl} = \Pi^{i}_{.jk(l)}, \tag{1.40}$$

$$\mathfrak{R}^{\cdots i}_{jkl} = R^{*\cdots i}_{jkl} - \frac{\delta_k{}^i}{n-1}R^{*\cdots h}_{hlj} - \frac{\delta_l{}^i}{n-1}R^{*\cdots h}_{khj},$$

where $R^{*\cdots i}_{jkl}$ denotes $R^{\cdots i}_{jkl}$ formed with the Π's in place of the Γ's.

8. The conditions

$$B^{\cdots i}_{jkl} = 0, \qquad R^{\cdots i}_{jkl} = 0 \tag{1.41}$$

are necessary and sufficient in order that it may be possible by transformation of coordinates to reduce the differential equations of the paths in the space to the form $x''^i = 0$, accordingly to reduce the curve length to

$$s = \int (A_i x''^i)^{1/p} dt.$$

If moreover $A_{i(j)(k)} = 0$, that is, if the A_i are linear homogeneous functions of x'^j,

$$A_{i(j)}x'^j = A_i, \qquad A_{i(j)} = -A_{j(i)}, \qquad p = 3.$$

The $A_{i(j)}$ are functions only of the x's. Then we have the curve length

$$s = \int (A_{i(j)}x''^i x'^j)^{1/3} dt.$$

For the two-dimensional affine space, $A_{i(j)}$ are all constants.

The conditions

$$\mathfrak{H}^{\cdots i}_{jkl} = 0, \qquad \mathfrak{R}^{\cdots i}_{jkl} = 0 \tag{1.42}$$

are necessary and sufficient in order that there may exist a coordinate system together with a parametrization such that the differential equations of the paths have the form $x''^i = 0$.

For the equivalence problem we have without difficulty the theorem:

THEOREM. *A necessary and sufficient condition for the equivalence of two spaces is that there exist an integer N ($\geqq 1$) such that the first N sets of following equations are compatible considered as equations for the determination of y^i and $P_j{}^i$ as functions of the independent variables x^i, and all their solutions*

$$y^i = y^i(x), \qquad P_j{}^i = \phi_j{}^i(x)$$

satisfy the $(N+1)$st set of the equations:

$$1.\quad \begin{cases} \overline{A}_i P_j^{\ i} = A_j, \\ \overline{B}_{jkl}^{\ \cdots i} P_r^{\ j} P_s^{\ k} P_t^{\ l} = B_{rst}^{\ \cdots u} P_u^{\ i}, \\ \overline{R}_{jkl}^{\ \cdots i} P_r^{\ j} P_s^{\ k} P_t^{\ l} = R_{rst}^{\ \cdots u} P_u^{\ i}, \end{cases}$$

$$2.\quad \begin{cases} \overline{\nabla}_j \overline{A}_i P_r^{\ j} P_s^{\ i} = \nabla_r A_s, \quad \overline{\nabla}'_j \overline{A}_i P_r^{\ j} P_s^{\ i} = \nabla'_r A_s, \\ \overline{\nabla}_h \overline{B}_{jkl}^{\ \cdots i} P_r^{\ h} P_s^{\ j} P_t^{\ k} P_u^{\ l} = \nabla_r B_{stu}^{\ \cdots v} P_v^{\ i}, \\ \overline{\nabla}'_h \overline{B}_{jkl}^{\ \cdots i} P_r^{\ h} P_s^{\ j} P_t^{\ k} P_u^{\ l} = \nabla'_r B_{stu}^{\ \cdots v} P_v^{\ i}, \\ \overline{\nabla}_h \overline{R}_{jkl}^{\ \cdots i} P_r^{\ h} P_s^{\ j} P_t^{\ k} P_u^{\ l} = \nabla_r R_{stu}^{\ \cdots v} P_v^{\ i}, \\ \overline{\nabla}'_h \overline{R}_{jkl}^{\ \cdots i} P_r^{\ h} P_s^{\ j} P_t^{\ k} P_u^{\ l} = \nabla'_r R_{stu}^{\ \cdots v} P_v^{\ i}, \\ \cdots\cdots\cdots\cdots\cdots\cdots. \end{cases}$$

Furthermore normal coordinates can be defined, and the replacement theorem also can be proved.

CHAPTER 2. THE CONNECTION C'

9. We shall now proceed to define another connection in our space. In a previous paper* the author has proved the theorem:

Let Φ_i be a covariant tensor of order m, that is, dependent upon x^i, $x'^i, \cdots, x^{(m)i}$; then n quantities

$$D_{ij}^{[m-\rho]}(\Phi)v^j = \sum_{\alpha=\rho}^{m} \binom{\alpha}{\rho} \frac{\partial \Phi_i}{\partial x^{(\alpha)j}} \frac{d^{\alpha-\rho} v^j}{dt^{\alpha-\rho}}$$

for a fixed positive integer ρ $(\leqq m)$ are the components of a covariant vector, provided that v^i is a contravariant vector.

In consequence of this theorem one can deduce from a vector v^i the following vectors:

$$(2.1)\quad -D_{ij}^{[1]}(T)v^j = 2G_{ij}\frac{dv^j}{dt} + G_{ik(l)}x''^k v^l + 2\Gamma_{i(l)}v^l,$$

$$(2.2)\quad D_{ij}^{[1]}(E)v^j = 3H_{ik}\frac{dv^k}{dt} + L_{ilk}x''^l v^k + M_{ik}v^k,$$

$$(2.3)\quad D_{ij}^{[2]}(E)v^j = 3H_{ik}\frac{d^2v^k}{dt^2} + 2(L_{ilk}x''^l + M_{ik})\frac{dv^k}{dt} + (H_{il(k)}x'''^l + H_{il(j)(k)}x''^l x''^j + M_{ij(k)}x''^j + P_{i(k)})v^k,$$

where

* A. Kawaguchi, *Some intrinsic derivations in a generalized space*, loc. cit.

$$\Gamma_i = G_{ik}\Gamma^k = \frac{1}{2(p-1)} M_{ik}x'^k, \tag{2.4}$$

$$H_{ik} = A_{i(k)} - A_{k(i)}, \qquad L_{ilk} = H_{il(k)} + H_{ik(l)}, \tag{2.5}$$

$$M_{ik} = A_{i(k)l}x'^l - A_{k(i)l}x'^l + A_{ki} + A_{ik} - B_{(i)(k)} + A_{ij(k)}x'^j, \tag{2.6}$$

$$P_i = A_{ijl}x'^j x'^l - B_{(i)l}x'^l + B_i. \tag{2.7}$$

It can be verified easily that the following relations hold:

$$\begin{aligned} -D^{[1]}_{ij}(T)x'^j &= (p-1)T_i, \\ -D^{[1]}_{ij}(E)x'^j &= (p-1)T_i, \\ D^{[2]}_{ij}(E)x'^j &= (p-1)E_i. \end{aligned} \tag{2.8}$$

10. Suppose now once more that $p \neq 3/2$. Then we have from (2.1) an absolute derivative along a curve

$$D^{[1]}_t v^i = \frac{dv^i}{dt} + G^{hi}(\tfrac{1}{2}G_{hk(l)}x''^k + \Gamma_{h(l)})v^l.$$

Since

$$G^{hi}(\tfrac{1}{2}G_{hk(l)}x''^k + \Gamma_{h(l)}) = \tfrac{1}{2}G^{hi}G_{hk(l)}(x''^k + 2\Gamma^k) + \Gamma^i_{(l)},$$

the absolute differential of a vector corresponding to a displacement from a line element (x, x') to a neighboring line element $(x+dx, x'+dx')$ is obtained without difficulty. We have

$$D^{[1]}v^i = dv^i + \Gamma^i_{(k)(j)}v^k dx^j + \tfrac{1}{2}G^{hi}G_{hj(k)}v^k \delta x'^j, \tag{2.9}$$

which reduces to (1.19), when $\delta x'^j = 0$.

The absolute differential (2.9) is, however, not invariant under transformation of t, because

$$G^{hi}G_{hk(l)}x'^k = (2p-3)G^{hi}A_{h(l)} - \delta^i_l \tag{2.10}$$

does not vanish in general. When and only when

$$(p-1)A_{i(k)} = A_{k(i)}, \tag{2.11}$$

(2.10) does vanish identically, and (2.9) is independent of transformation of t.

11. If we adopt the notations

$$\begin{aligned} I_{lji} &= A_{l(j)(i)} + A_{j(l)(i)} + (p-3)A_{i(j)(l)}, \\ J_{li} &= A_{l(i)} + (p-2)A_{i(l)}; \end{aligned} \tag{2.12}$$

it will be easily inferred from (1.4) and (1.5) that

$$I_{lji}x'^i = 0, \qquad J_{li}x'^i = 0, \qquad I_{lji}x'^j = (p-4)J_{li}.$$

Hence the tensor

$$\tilde{C}^{k}_{\cdot ij} = (p-3)G^{lk}I_{lji} + G^{lk}_{\cdot\cdot(j)}J_{li} \tag{2.13}$$

satisfies the relations

$$\tilde{C}^{k}_{\cdot ij}x'^{i} = 0, \qquad \tilde{C}^{k}_{\cdot ij}x'^{j} = 0. \tag{2.14}$$

We shall now put

$$\begin{aligned} C^{k}_{\cdot ij} &= \frac{1}{(p-3)^2}\tilde{C}^{k}_{\cdot ij}, \qquad p \neq 3, \\ &= \tilde{C}^{k}_{\cdot ij}, \qquad p = 3. \end{aligned} \tag{2.15}$$

The tensor $C^{k}_{\cdot ij}$ is constructed from only A_i and is independent of B. With the help of the tensor $C^{k}_{\cdot ij}$ an absolute differential can be defined

$$\delta v^{i} = dv^{i} + \Gamma^{i}_{(j)(k)}v^{j}dx^{k} + C^{i}_{\cdot jk}v^{j}\delta x'^{k}, \tag{2.16}$$

which is invariant under transformation of the parameter t. The connection characterized by A_i, Γ^{i}, and $C^{i}_{\cdot jk}$ will be referred to as the connection C'. The torsion* of this connection is $C^{i}_{\cdot jk}$, and for $v^{i}=x'^{i}$ (2.16) gives us the base connection (1.20).

12. The covariant derivatives referred to the connection C' possess the form

$$\nabla_j v^{i} = \frac{\partial v^{i}}{\partial x^{j}} - \frac{\partial v^{i}}{\partial x'^{k}}\Gamma^{k}_{(j)} + \Gamma^{i}_{(k)(j)}v^{k}, \quad \tilde{\nabla}'_j v^{i} = \frac{\partial v^{i}}{\partial x'^{j}} + C^{i}_{\cdot kj}v^{k}, \tag{2.17}$$

and we have

$$\begin{aligned} (\nabla_j\nabla_k - \nabla_k\nabla_j)v^{i} &= -\tilde{R}^{\cdot\cdot\cdot i}_{jkl}v^{l} + K^{\cdot\cdot l}_{jk}\tilde{\nabla}'_l v^{i}, \\ (\nabla_j\tilde{\nabla}'_k - \tilde{\nabla}'_k\nabla_j)v^{i} &= -\tilde{B}^{\cdot\cdot\cdot i}_{jkl}v^{l} + C^{l}_{\cdot jk}\nabla_l v^{i}, \\ (\tilde{\nabla}'_j\tilde{\nabla}'_k - \tilde{\nabla}'_k\tilde{\nabla}'_j)v^{i} &= -P^{\cdot\cdot\cdot i}_{jkl}v^{l} - 2C^{h}_{\cdot[kj]}\tilde{\nabla}'_h v^{i}, \end{aligned} \tag{2.18}$$

where

$$\begin{aligned} \tilde{R}^{\cdot\cdot\cdot i}_{jkl} &= R^{\cdot\cdot\cdot i}_{jkl} + K^{\cdot\cdot h}_{jk}C^{i}_{\cdot lh}, \\ \tilde{B}^{\cdot\cdot\cdot i}_{jkl} &= \Gamma^{i}_{(l)(j)(k)} + C^{i}_{\cdot hk}\Gamma^{h}_{(l)(j)} - \frac{\partial C^{i}_{\cdot lk}}{\partial x^{j}} + C^{i}_{\cdot l(kh)}\Gamma^{h}_{(j)} \\ &\quad - \Gamma^{i}_{(h)(j)}C^{h}_{\cdot lk} + \Gamma^{h}_{(j)(k)}C^{i}_{\cdot lh}, \\ P^{\cdot\cdot\cdot i}_{jkl} &= C^{i}_{\cdot lj(k)} - C^{i}_{\cdot lk(j)} + C^{i}_{\cdot hk}C^{h}_{\cdot lj} - C^{i}_{\cdot hj}C^{h}_{\cdot lk} \end{aligned} \tag{2.19}$$

are the curvature tensors.

* See Cartan, *Les espaces de Finsler*, loc. cit., p. 32.

Consider the differential forms

$$\begin{aligned} dx^i, \qquad \omega^i &= dx'^i + \Gamma^i_{(j)} dx^j, \\ \omega^i_j &= \Gamma^i_{(j)(k)} dx^k + C^i_{.jk}\omega^k; \end{aligned} \tag{2.20}$$

then we have the equations of structure* of the connection C':

$$\begin{aligned} [dx^j \omega^i_j] &= C^i_{.jk}[dx^j \omega^k], \\ (\omega^i)' + [\omega^i_j \omega^j] &= -\tfrac{1}{2} K^{\cdots i}_{jk}[dx^j dx^k] - C^i_{.[jk]}[\omega^j \omega^k], \\ (\omega^i_j)' + [\omega^i_k \omega^k_j] &= -\tfrac{1}{2}\bar{R}^{\cdots i}_{klj}[dx^k dx^l] - B^{\cdots i}_{klj}[dx^k \omega^l] - \tfrac{1}{2} P^{\cdots i}_{klj}[\omega^k \omega^l]. \end{aligned} \tag{2.21}$$

The identities of Bianchi are now expressed by

$$\begin{aligned} \nabla_{[h}\bar{R}^{\cdots i}_{jk]l} + K^{\cdots r}_{[hj}\bar{B}^{\cdots i}_{k]lr} &= 0, \\ 2\nabla_{[h}\bar{B}^{\cdots i}_{j]kl} + \bar{\nabla}'_l R^{\cdots i}_{hjk} + P^{\cdots i}_{lrk} K^{\cdots r}_{hj} &= 0, \\ 2\bar{\nabla}'_{[m}\bar{B}^{\cdots i}_{|k|l]j} + \nabla_k P^{\cdots i}_{lmj} &= 0, \\ \bar{\nabla}'_{[h} P^{\cdots i}_{jk]l} &= 0. \end{aligned} \tag{2.22}$$

The equivalence theorem can be proved, but we shall forego the details here.

---

* See Cartan, loc. cit., pp. 32–36.

HOKKAIDO IMPERIAL UNIVERSITY,
SAPPORO, JAPAN

ON THE REDUCTION OF DYNAMICAL SYSTEMS BY MEANS OF PARAMETRIZED INVARIANT RELATIONS*

BY

E. R. VAN KAMPEN AND AUREL WINTNER

Introduction. The classical reduction theory of canonical systems with n degrees of freedom assumes that there is known a function group of first integrals.† The reduction of the degree of freedom is then carried out by using the Hamilton-Jacobi theory or an equivalent approach. In the present paper a more general problem will be considered, since it will not be assumed that the known functions, or hypersurfaces in the phase-space, are represented by first integrals. In fact, the first integrals will be replaced by invariant relations,‡ so that, in particular, the known hypersurfaces need not form a continuous family, but may be isolated.

It should be mentioned that the generalization of the reduction problem to the case where the given first integrals are replaced by invariant relations is not an artificial problem but one which arises quite naturally in the simplest applications. If, for instance, one wants to reduce the degree of freedom of the problem of three bodies by means of the classical first integrals in an explicit and symmetrical manner, one is compelled to replace these first integrals by certain of their combinations which form a complicated system of invariant relations and not a system of first integrals.§

The treatment of the general case of invariant relations will differ from the usual treatment of the case of first integrals in that all considerations will be based on a parametrization of the system of invariant relations, this parametrization being symmetrical with respect to the n coordinates and n impulses. Needless to say, the usual treatment of first integrals, based on the Hamilton-Jacobi theory, cannot be applied in the general case under consideration.

In the particular case where the system of invariant relations is a system of first integrals (§10), the treatment of the general case goes over into a

* Presented to the Society, April 16, 1938, under the title, *The canoncical reduction of Hamilton systems*; received by the editors August 2, 1937.

† A presentation of the classical reduction theory is given, for instance, by Engel [1].

‡ As to this notion, which is due to Jacobi and is fundamental for the geometrical theories of Poincaré and Birkhoff, cf. Levi-Civita [4].

§ van Kampen and Wintner [3].

reduction theory which seems to have essential formal advantages over the usual treatments of this particular case.

In the limiting case where the system of invariant relations is empty, the considerations go over into a treatment of canonical transformations which has been recently developed* and which suggested the approach of the present paper to the reduction problem.

§11, which is independent of the rest of the paper, attempts an application of the method of parametrization to the case where the given relations are not invariant systems but constraints. However, the usual rule for the introduction of constraints will be verified to be identical with the rule suggested by the method of parametrization only in the case where the constraints are holonomic.†

Notation. A matrix with k rows and l columns will be called a (k, l)-matrix, so that the product CD of a (k, l)-matrix C and an (l, m)-matrix D is a (k, m)-matrix. If C is a (k, l)-matrix, C' will denote the (l, k)-matrix which is the transposed of C. A $(k, 1)$-matrix will be termed a k-vector, so that, if C and D are k-vectors, their scalar product may be represented as either of the two matrix products $C'D$, $D'C$. The (k, l)-matrix in which all kl elements are 0 will be denoted by $0^k{}_l$; and $0^k{}_1$, by 0^k. The letters $k, r, \cdots$ will denote scalar functions. Total differentiation with respect to the time t will be marked by a dot, while t as a subscript will refer to partial differentiation. If, for instance, c is a scalar function $c(Z, t)$ of a k-vector Z and of t, and if Z is a function $Z(t)$ of t, then

$$\dot{c} = c_t + \dot{Z}' \operatorname{grad}_Z c,$$

where the subscript of grad refers to the space in which one carries out the partial differentiations. The functions in question will be assumed to possess, in the regions under consideration, continuous derivatives of the μth order, where either $\mu = 1$ or $\mu = 2$ or $\mu = 3$; it will be always clear from the connection which of these three values is the actual value of μ.

1. **Parametrizations of invariant systems.** For a fixed integer $n \geqq 1$, let G denote the $(2n, 2n)$-matrix

$$G = \begin{pmatrix} 0^n{}_n & -E_n \\ E_n & 0^n{}_n \end{pmatrix}, \tag{1}$$

where E_n denotes the n-rowed unit matrix, so that

$$G = -G' = -G^{-1}. \tag{2}$$

* Wintner [7], van Kampen and Wintner [2].

† As to the holonomic case, cf. Levi-Civita [5].

The canonical system with n degrees of freedom and with the Hamiltonian function

$$h = h(q_1, \cdots, q_n, p_1, \cdots, p_n, t),$$

that is, the system

$$\dot{q}_\lambda = \frac{\partial h}{\partial p_\lambda}, \qquad \dot{p}_\lambda = -\frac{\partial h}{\partial q_\lambda}, \qquad \lambda = 1, \cdots, n,$$

may clearly be written in the form

$$G\dot{X} = \operatorname{grad}_X h, \tag{3}$$

where X denotes the $2n$-vector

$$X = (x_1, \cdots, x_{2n}) \tag{4}$$

defined by

$$X = \begin{pmatrix} Q \\ P \end{pmatrix}, \qquad Q = \begin{pmatrix} q_1 \\ \cdot \\ \cdot \\ \cdot \\ q_n \end{pmatrix} \equiv \begin{pmatrix} x_1 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{pmatrix}, \qquad P = \begin{pmatrix} p_1 \\ \cdot \\ \cdot \\ \cdot \\ p_n \end{pmatrix} \equiv \begin{pmatrix} x_{n+1} \\ \cdot \\ \cdot \\ \cdot \\ x_{2n} \end{pmatrix} \tag{4 bis}$$

so that

$$h = h(X, t). \tag{5}$$

For a fixed m, which is independent of (X, t) and is such that

$$m \leqq 2n, \tag{6}$$

let K be a $(2n-m)$-vector defined as a function $K(X, t)$, in the $(2n+1)$-dimensional (X, t)-region under consideration, in such a way that

$$2n - m = \text{rank of } \operatorname{grad}_X K, \tag{7}$$

and that the condition

$$K \equiv K(X, t) = 0^{2n-m} \tag{8}$$

determines a non-empty set $\Lambda(t)$ in the $2n$-dimensional X-space for every fixed t. Since $\operatorname{grad}_X K$ is a $(2n-m, 2n)$-matrix, the assumption (7) is to the effect that $\Lambda(t)$ is a locally m-dimensional manifold for every fixed t. Hence one can choose a $2n$-vector

$$F = F(Y, t), \tag{9}$$

depending on t and on an m-vector

$$Y = (y_1, \cdots, y_m) \tag{10}$$

in such a way that

$$X = F(Y, t) \tag{11}$$

is a local parametrization of the manifold $\Lambda(t)$, and that

$$m = \text{rank of } J, \tag{12}$$

where J denotes the Jacobian $(2n, m)$-matrix

$$J = J(Y, t) = \text{grad}_Y F. \tag{13}$$

If $Y = Y(t)$ is an arbitrary curve in the space of the m parameters (10), the corresponding curve

$$X(t) = F(Y(t), t) \tag{14}$$

in the space of the $2n$ phase coordinates (4) is such that the relation

$$\dot{X} = F_t + J\dot{Y} \tag{15}$$

is an identity in t.

The system (8) of $2n-m$ relations is called an *invariant system of* (3) if every solution $X(t)$ of (3) which is on the manifold $\Lambda(t)$ for a single t is on $\Lambda(t)$ for every t, where the manifold $\Lambda(t)$ is supposed to be a non-empty set. Any parametrization (11) of an invariant system (8) of (3) will also be called an invariant system of (3). An invariant system (8) of (3) does not determine the function (9) uniquely, since (12) is the only restriction as to the choice of the parameters (10). In what follows, the parametrization (11) of (8) will be considered as fixed.

It is clear from (12) that if a given solution $X(t)$ of (3) satisfies the invariant system (8), then there exists in the space of the parameter vector Y exactly one path $Y = Y(t)$ for which (14) is valid. According to (15) and (3), this unique $Y(t)$ is such that the relation

$$GJ\dot{Y} = -GF_t + \text{grad}_X\, h \tag{16}$$

is, in virtue of (11), an identity in t. Now there exists for every point of the manifold $\Lambda(t)$ exactly one solution $X(t)$ of (3) passing through this point. Hence (16) and (12) imply that there belongs to the parametrization (11) of the invariant system (8) of (3) a unique m-vector

$$L = L(Y, t) \tag{17}$$

which is defined in the $(m+1)$-dimensional (Y, t)-region in such a way that the relation

$$GJL = -GF_t + \text{grad}_X\, h \tag{18}$$

is, in virtue of (11), an identity in Y and t. Since both $Y(t)$ and $L(Y, t)$ are uniquely determined, comparison of (16) with (18) shows that

$$\dot{Y} = L(Y, t). \tag{19}$$

Accordingly, $X(t)$ is a solution of (3) satisfying the invariant system (8) of (3) if and only if the system (19) of m ordinary differential equations has a solution $Y(t)$ by means of which $X(t)$ is representable in the form (14). It is clear, from the proof, that the two differential equations (19), (16) are equivalent, although (19) has, and (16) need not have, the normal form of ordinary differential equations.

2. **Some formal consequences of the properties of G.** On defining an m-vector R by

$$R = R(Y, t) = \tfrac{1}{2}J'GF \tag{20.1}$$

and a scalar r by

$$r = r(Y, t) = \tfrac{1}{2}F_t'GF, \tag{20.2}$$

one will be able to show that the matrix function $J'GJ$ of (Y, t) may be represented in the (Y, t)-region under consideration as the alternating derivative (curl)

$$J'GJ = (\mathrm{grad}_Y R) - (\mathrm{grad}_Y R)', \tag{21.1}$$

while the vector function JGF_t of (Y, t) appears in the form

$$J'GF_t = R_t - \mathrm{grad}_Y r. \tag{21.2}$$

In fact, (21.1) and (21.2) are the identities

$$2J'GJ = \mathrm{grad}_Y (J'GF) - (\mathrm{grad}_Y (J'GF))'$$

and

$$2J'GF_t = (J'GF)_t - (\mathrm{grad}_Y (F_t'GF))'$$

which follow from (2) and (13), since the second derivatives cancel.

Since the Jacobian (13) is a $(2n, m)$-matrix, it is clear from (1) that $J'GJ$ is a skew-symmetric (m, m)-matrix, so that

$$2l = \text{rank of } J'GJ \tag{22}$$

for some integer l which satisfies the inequality

$$2l \leqq m \tag{23}$$

and will be supposed to be independent of (Y, t) in the $(m+1)$-dimensional

(Y, t)-region under consideration. It may be mentioned that (6) and (25) can be replaced by the sharper statement that

$$0 \leqq 2l \leqq m \leqq n + l \leqq 2n. \tag{23 bis}$$

While the definition (23) of the integer l assumes, in view of (13), the choice of a parametrization (11) of the invariant system (8), it will be shown that l is uniquely determined by the invariant system (8) alone. This is implied by the relation

$$2m - \text{rank of } J'GJ = 2n - \text{rank of } NGN', \tag{24}$$

where N denotes the $(2n-m, 2n)$-matrix

$$N = N(X, t) = \text{grad}_X K, \tag{25}$$

so that NGN' is, in view of (1), a skew-symmetric $(2n-m, 2n-m)$-matrix. In fact, since (11) is a parametrization of (8), $K(F(Y, t), t) = 0^{2n-m}$ is an identity in (Y, t). On differentiating this identity with respect to the m components of Y, one sees from (26) and (13) that

$$NJ = 0^{2n-m} \tag{26.1}$$

is an identity in (Y, t) in virtue of (11). Furthermore, it is seen from (12) that there exists an $(m, 2n)$-matrix $T = T(Y, t)$ such that

$$TJ = E_m, \tag{26.2}$$

where E_m is the m-rowed unit matrix. Now it is clear from (7), (25), (26.1) and (26.2) that the $(2n, 2n)$-matrix

$$\begin{pmatrix} N \\ T \end{pmatrix}$$

has the rank $2n$; that is, it has a non-vanishing determinant; and its reciprocal matrix is, by (26.1) and (26.2), of the form

$$\begin{pmatrix} N \\ T \end{pmatrix}^{-1} = (WJ),$$

where $W = W(Y, t)$ is some $(2n, 2n-m)$-matrix. Hence (2) implies that the two $(2n, 2n)$-matrices

$$H_1 = \begin{pmatrix} N \\ T \end{pmatrix} G \begin{pmatrix} N \\ T \end{pmatrix}' = \begin{pmatrix} NGN' & NGT' \\ TGN' & TGT' \end{pmatrix}$$

and

$$H_2 = (WJ)'G(WJ) = \begin{pmatrix} W'GW & W'GJ \\ J'GW & J'GJ \end{pmatrix}$$

are reciprocal matrices. Now a well known theorem states that, since H_1 and H_2 are reciprocal matrices, the difference of order and rank in NGN' is the same as in $J'GJ$. This proves (24). A direct proof of (24) follows by multiplication of H_1 and the matrix

$$\begin{pmatrix} E_{2n-m} & W'GJ \\ 0^m{}_{2n-m} & J'GJ \end{pmatrix}.$$

The Lagrange bracket

$$[u, v] = \sum_{\lambda=1}^{n} \frac{\partial(a_{n+\lambda}, a_\lambda)}{\partial(u, v)} = -[u, v] \tag{27.1}$$

of a $2n$-vector

$$(a_1, \cdots, a_{2n}) = A = A(u, v),$$

which depends on two scalar parameters u, v, may be written in the form

$$[u, v] = A_u' G A_v, \tag{27.2}$$

so that the well known identity

$$[u, v]_w + [v, w]_u + [w, u]_v = 0, \tag{27.3}$$

which holds for an arbitrary $2n$-vector $A = A(u, v, w)$ depending on three scalar parameters, may be written as

$$(A_u' G A_v)_w + (A_v' G A_w)_u + (A_w' G A_u)_v = 0. \tag{27.4}$$

Similarly, the Poisson parenthesis

$$(f_1, f_2) = \sum_{\lambda=1}^{n} \frac{\partial(f_1, f_2)}{\partial(x_{n+\lambda}, x_\lambda)} = -(f_2, f_1) \tag{28.1}$$

of two scalars

$$f_1 = f_1(X, t), \qquad f_2 = f_2(X, t)$$

may be written as

$$(f_1, f_2) = (\mathrm{grad}_X f_1)' G (\mathrm{grad}_X f_2). \tag{28.2}$$

It may be mentioned that the matrix function $J'GJ$ of (Y, t) is independent of t if and only if the vector function $J'GF_t$ of (Y, t) is the Y-gradient of a scalar function of (Y, t). In fact, $(J'GJ)_t$ is the alternating derivative (curl) of $J'GF_t$ with respect to Y. In order to verify this, it is sufficient to replace in (27.4) the variables A, u, v, w by F, t and two components of Y, respectively.

3. **Separating parametrizations of invariant systems.** It will be convenient to split the m-vector (10), with the use of the integer (22) satisfying (23), into the $2l$-vector.

$$Y^0 = (y_1, \cdots, y_{2l}) \tag{29.1}$$

and the $(m-2l)$-vector

$$Y^* = (y_{2l+1}, \cdots, y_m), \tag{29.2}$$

so that

$$(y_1, \cdots, y_m) = Y = \begin{pmatrix} Y^0 \\ Y^* \end{pmatrix}. \tag{29.3}$$

The limiting case

$$2l = m, \quad \text{that is,} \quad Y = Y^0, \tag{29 bis}$$

will not be excluded.

The parametrization (11) of (8) will be called a *separating parametrization* if there exists a $(2l, 2l)$-matrix $S^0 = S^0(Y, t)$ by means of which the (m, m)-matrix (21.1) may be represented in the form

$$J'GJ = \begin{pmatrix} S^0 & 0^{2l}{}_{m-2l} \\ 0^{m-2l}{}_{2l} & 0^{m-2l}{}_{m-2l} \end{pmatrix}. \tag{30}$$

It is clear from (22) and from the skew-symmetry of (21.1) that

$$\det S^0 \neq 0 \tag{31}$$

and

$$S^0 = -(S^0)', \tag{32}$$

whenever S^0 exists.

It will be shown that if $S^0 = S^0(Y, t)$ exists, that is, if (11) is a separating parametrization of the invariant system (8) of (3), then there exist a $2l$-vector

$$R^0 = R^0(Y^0, t) \tag{33.1}$$

and a scalar

$$h^0 = h^0(Y^0, t), \tag{33.2}$$

such that the $(2l, 2l)$-matrix S^0 occurring in (30) may be represented as the alternating derivative

$$S^0(Y^0, t) \equiv S^0 = (\text{grad}_{Y^0} R^0) - (\text{grad}_{Y^0} R^0)', \tag{34.1}$$

while the m-vector (21.2) appears in the form

$$J'GF_t = \operatorname{grad}_Y h + \begin{pmatrix} R^0{}_t - \operatorname{grad}_{Y^0} h^0 \\ 0^{m-2l} \end{pmatrix}, \tag{34.2}$$

where

$$h = h(Y; t) \equiv h(F(Y, t), t), \tag{35}$$

cf. (5), (11). This implies, in particular, that in case of a separating parametrization the functions

$$J'GJ,\ J'GF_t - \operatorname{grad}_Y h$$

of t and of the m-vector (29.3) are both independent of the $(m-2l)$-vector (29.2). In the proof use will be made of the identity

$$J' \operatorname{grad}_X h = \operatorname{grad}_Y h, \tag{36}$$

which is obvious from (11), (13), and (35).

First, it is clear that the vector function R is determined by the conditions (21.1), (21.2) only up to an additive term of the form

$$\operatorname{grad}_Y b, \tag{37}$$

where $b = b(Y, t)$ is an arbitrary scalar function. If (37) is added to R, then the term b_t has to be added to r in order to satisfy (21.2). On the other hand, on placing

$$R = \begin{pmatrix} R^0 \\ R^* \end{pmatrix}, \tag{38}$$

so that $R^0 = R^0(Y, t)$ and $R^* = R^*(Y, t)$ are the $2l$-vector and the $(m-2l)$-vector formed by the first $2l$ and the last $m-2l$ components of the m-vector R, one can see from (29.3) that (21.1) implies the pair of identities

$$(\operatorname{grad}_{Y^0} R^0) - (\operatorname{grad}_{Y^0} R^*)' = 0^{2l}{}_{m-2l}, \tag{39.1}$$

and

$$(\operatorname{grad}_{Y^*} R^*) - (\operatorname{grad}_{Y^*} R^*)' = 0^{m-2l}{}_{m-2l}, \tag{39.2}$$

if the condition (30) for a separating parametrization is satisfied. Since (39.2) is precisely the integrability condition for the existence of a scalar

$$b = b(Y, t) = b(Y^0, Y^*, t) \tag{40}$$

such that

$$R^* = \operatorname{grad}_{Y^*} b \tag{41}$$

is an identity in $(Y, t) = (Y^0, Y^*, t)$, and since (38) is undetermined up to an additive term of the form (37), it is clear that, on modifying the functions R and r in a suitable manner, one can assume without loss of generality that the function (40) is identically zero. Then $R^* = 0^{m-2l}$ by (41); hence

$$R = \begin{pmatrix} R^0 \\ 0^{m-2l} \end{pmatrix} \tag{42}$$

by (38), and (39.1) goes over into

$$\operatorname{grad}_{Y^*} R^0 = 0^{2l}{}_{m-2l}. \tag{42 bis}$$

Now it is seen from (42 bis) and (29.3) that the function $R^0 = R^0(Y, t)$ is of the form (33.1).

In order to prove the existence of the function (33.2), notice first that on multiplying (18) by $-J'$ and using (36), one obtains

$$-J'GJL = J'GF_t - \operatorname{grad}_Y h. \tag{43}$$

Since substitution of (42) into (21.2) gives

$$J'GF_t - \operatorname{grad}_Y h = \begin{pmatrix} R^0{}_t - \operatorname{grad}_{Y^0}(r + h) \\ 0^{m-2l} - \operatorname{grad}_{Y^*}(r + h) \end{pmatrix}, \tag{44}$$

it is seen from (30) and (43) that

$$0^{m-2l} = \operatorname{grad}_{Y^*}(r + h). \tag{45}$$

Now (45) means that the function $r+h$ of $(Y, t) = (Y^0, Y^*, t)$ is independent of Y^*; that is, that the scalar function h^0 defined by

$$h^0 = r + h \tag{46}$$

is of the form (33.2). Finally, substitution of (45) and (46) into (44) gives (34.2), while (34.1) follows from (30) if one substitutes (33.1) and (42) into (21.1).

The theorem thus proved and the last remark in §2 imply that in the separating case characterized by (30) the function (34.1) of (Y^0, t) is independent of t if and only if the function (33.1) is also, or, more precisely, that

$$S^0{}_t = 0^{2l}{}_{2l} \quad \text{if and only if} \quad R^0{}_t \equiv 0^{2l} \tag{47}$$

holds for a *suitable* choice of the function (33.1).

4. **The separated differential equations belonging to a separating parametrization of an invariant system.** The name "separating parametrization" will now be justified by showing that if the parametrization (11) of an invari-

ant system (8) of (3) is a separating parametrization, then the first $2l$ of the m differential equations represented by (19) appear in the form

$$\dot{Y}^0 = L^0(Y^0, t), \tag{48}$$

that is, as a system of $2l$ differential equations which is separated from the $m-2l$ differential equations determining the components (29.2) of the m-vector (29.3) as functions of t. The latter differential equations are, in view of (19) and (29.3), of the form

$$\dot{Y}^* = L^*(Y^*, Y^0, t). \tag{49}$$

Thus if there is known a solution $Y^0(t)$ of (48), the corresponding functions $Y^*(t)$ are the solutions of

$$\dot{Y}^* = L^*(Y^*, t), \tag{50}$$

where

$$L^*(Y^*, t) = L^*(Y^*, Y^0(t), t) \tag{50 bis}$$

(cf. (49)) is a known function of t and of the $(m-2l)$-vector (29.2), so that (19) splits into the two separated systems (48), (50). Finally, the m-vector $Y(t)$, determined by $Y^0(t)$ and $Y^*(t)$ in virtue of (29.3), gives the corresponding solution (14) of (3). The system (48) of order (22) will be referred to as the separated system belonging to the separating parametrization (11) of (8).

In order to prove the possibility of splitting (19) into (48) and (50), notice first that (19) is, as pointed out at the end of §1, equivalent to the system (16). Now, if (16) is multiplied by J', the first $2l$ of the resulting m differential equations appear, in view of (30), (34.2), (36), and (29.3), as

$$S^0\dot{Y}^0 = -R^0{}_t + \operatorname{grad}_{Y^0} h^0. \tag{51}$$

This completes the proof, since (33.1), (33.2), (34.1), and (31) imply that (51) can be written in the form (48). Needless to say, (48) is in the limiting case (29 bis) identical with the original system (19). The result just proved contains, however, information for this limiting case also.

In fact, whether (29 bis) is or is not satisfied, the result proved above may be described by saying that if the parametrization (11) of an invariant system (8) of (3) is a separating parametrization, then the $2l$-vector (29.2) is determined by a system of $2l$ differential equations which form a *non-degenerate Pfaffian dynamical system* with l degrees of freedom.† In order to see this, it

† The term *Pfaffian dynamical system* is used in the sense of Birkhoff. Such a system is called non-degenerate if it can be solved with respect to the time derivatives, that is, if the skew-symmetric matrix, represented by the curl of the vector potential involved, does not vanish in the domain of the phase space which is under consideration.

is sufficient to notice that on the one hand (34.1) and (51) are equivalent to

$$\{(\text{grad}_{Y^0} R^0) - (\text{grad}_{Y^0} R^0)'\}\dot{Y}^0 = -R^0{}_t + \text{grad}_{Y^0} h^0, \tag{52}$$

and that on the other hand (52) clearly is the system of Euler-Lagrange equations which belongs to the calculus of variations problem

$$\delta \int \{(\dot{Y}^0)' R^0(Y^0, t) + h^0(Y^0, t)\} dt = 0, \tag{52 bis}$$

where the ends are fixed and the problem is non-degenerate in view of (31) and (34.1).

5. **Existence of separating parametrizations of an invariant system.** The considerations of §3 and §4 assumed that there is *given* a separating parametrization (11) of the invariant system (8) of (3). Actually, one can always *choose* the parametrization of an invariant system as a separating parametrization. This existence theorem can be inferred from a classical theorem in the theory of Pfaffians as follows.

First, (21.1), (21.2), (36), and (16) hold for arbitrary parametrizations of (8). Let (11) be a given parametrization. Since from (16) and (36)

$$J'GJ\dot{Y} + J'GF_t - \text{grad}_Y h = 0^m, \tag{53}$$

it is seen from (21.1) and (21.2) that

$$\{(\text{grad}_Y R) - (\text{grad}_Y R)'\}\dot{Y} + R_t - \text{grad}_Y (r + h) = 0^m, \tag{54}$$

where R, r, and h are the functions (20.1), (20.2), and (35). If one denotes by ω the $(m+1)$-ary Pfaffian

$$\omega = R'dY + (r + h)dt \tag{55}$$

(cf. (10)), it is clear that (53) is the system of Euler-Lagrange equations belonging to

$$\delta \int \omega \equiv \delta \int \{[R(Y, t)]'dY + [r(Y, t) + h(Y, t)]dt\} = 0, \tag{56}$$

where the ends are fixed. Notice that (56), in contrast with (52 bis), need not be non-degenerate, since the determinant of the skew-symmetric (m, m)-matrix (21.1) vanishes identically unless (29 bis) is satisfied.

Let G^0 denote the $(2l, 2l)$-matrix

$$G^0 = \begin{pmatrix} 0^l{}_l & -E_l \\ E_l & 0^l{}_l \end{pmatrix} = -(G^0) = -(G^0)^{-1}, \tag{57}$$

obtained from (1), (2) by replacing n by l, where l is defined by (22) and satis-

fies (23 bis). It is well known† that on subjecting the m-dimensional Y-space to a suitable non-singular local point transformation for every fixed t, and omitting then an additive complete differential, one can transform the Pfaffian (55) into the $(2l+1)$-ary normal form

$$-\tfrac{1}{2}Y^{0\prime}G^0dY^0 + h^0dt, \tag{58}$$

it being understood that in (58) the notation (29.3) is applied to the *new* m-vector Y and that

$$h^0 = h^0(Y^0, Y^*, t) \tag{59}$$

is some scalar function. Since a variation problem (56) is covariant under point transformations, and since a complete differential can always be omitted beneath the signs $\delta\int$, where the ends are fixed, it is clear from (59) and from the normal form (58) of (55) that (56) appears in terms of the new m-vector Y in the form

$$\delta\int\{-\tfrac{1}{2}Y^{0\prime}G^0dY^0 + h^0(Y^0, Y^*, t)dt\} = 0. \tag{60}$$

Now, (57) being a constant matrix, it is seen from (29.3) that the Euler-Lagrange equations belonging to (60) are

$$\begin{pmatrix} G^0 & 0^{2l}{}_{m-2l} \\ 0^{m-2l}{}_{2l} & 0^{m-2l}{}_{m-2l} \end{pmatrix}\begin{pmatrix} \dot{Y}^0 \\ \dot{Y}^* \end{pmatrix} - \begin{pmatrix} \operatorname{grad}_{Y^0} h^0 \\ \operatorname{grad}_{Y^*} h^0 \end{pmatrix} = 0^m. \tag{61}$$

On substituting (29.3) into (57) and comparing the result with (61), one obtains

$$J'GJ = \begin{pmatrix} G^0 & 0^{2l}{}_{m-2l} \\ 0^{m-2l}{}_{2l} & 0^{m-2l}{}_{m-2l} \end{pmatrix}. \tag{62}$$

Consequently, the condition (30), which characterizes the separating parametrizations, is satisfied, S^0 being the matrix (57). This completes the proof of the existence theorem stated at the beginning of this section.

Let it be mentioned for application in §6 that the function (59) occurring in (61) can be chosen, in view of the results of §3, as a function of the form (33.2). In other words, one can write (61) in the form

$$G^0\dot{Y}^0 = \operatorname{grad}_{Y^0} h^0 \tag{63}$$

in view of $\operatorname{grad}_{Y^*} h^0 \equiv 0^{m-2l}$.

† The normal form usually given in textbooks is not (58) but another Pfaffian. However, the latter differs from the Pfaffian (58) only in a complete differential; cf., e.g., van Kampen and Wintner [2], p. 862.

6. **Canonical parametrizations of an invariant system.** A given parametrization (11) of a system (8) of m invariant relations of a canonical system (3) with n degrees of freedom will be called a *canonical parametrization* if it is a separating parametrization (§3), and the corresponding separated system of $2l$ differential equations, that is, the system (48), is a canonical system with l degrees of freedom, so that (52) is of the particular form (63). Thus the result found at the end of §5 may be expressed by saying that the invariant system (8) of (3) always admits a parametrization which is a canonical parametrization.

The construction of a canonical parametrization of a given invariant system depends on the construction of a point transformation which transforms the Pfaffian (55) into its normal form (58). And the construction of such a point transformation is known to require the solution of a complete system of partial differential equations, a solution which, in general, cannot be obtained by mere quadratures. Hence, if one does not want to refer to the existence theorem of complete systems, which depends on successive approximations or on equivalent processes, one has to assume that the point transformation in question is a priori known. This is the situation in many applications where the geometrical or dynamical connection suggests the "correct" choice of the parameters (10). Actually, it is easy to see from the considerations of §5 that the point transformation which transforms the Pfaffian (55) into its normal form (58) can be chosen such as to depend only on the function (8) and not on the function (5). In other words, (8) admits a parametrization (11) which is a *fixed* canonical parametrization for *all* those canonical systems (3), with n degrees of freedom, for which the *fixed* system (8) of m relations is an invariant system.

It will be assumed in what follows that both (8) and (3) are fixed. The considerations of §5 clearly imply that if a given separating parametrization (11) of (8) is such that (30) is of the particular form (62), that is, if

$$S^0 = G^0 \tag{64}$$

holds for the constant $(2l, 2l)$-matrix (57) and for the matrix (34.1) occurring in the definition (30) of a separating parametrization, then the given separating parametrization is a canonical parametrization.

If a canonical parametrization (11) of an invariant system (8) of (3) is such that, for some constant m-vector C, the Y_0-gradient of the Hamiltonian function (33.2) of the separated system (63) is the 0^m-vector for every t, then

$$Y^0(t) \equiv C$$

clearly is a solution of (63). For this solution of equilibrium, (50) goes over, in view of (50), into

$$\dot{Y}^* = L^*(Y^*, C, t), \tag{65}$$

so that

$$Y(t) = \begin{pmatrix} Y^0(t) \\ Y^*(t) \end{pmatrix} \equiv \begin{pmatrix} C \\ Y^*(t) \end{pmatrix} \tag{66}$$

is, for every solution $Y^*(t)$ of (65), a solution of (19). The solutions $X(t)$ of (3) obtained from (66.1) by means of (14) are, in the main, the stationary solutions of Routh and Levi-Civita.†

7. **Completely canonical parametrizations.** A canonical parametrization (11) of an invariant system (8) of (3) will be said to be *completely canonical* if the Hamiltonian function (33.2) of the separated canonical system (63) may be obtained by direct substitution of (11) and (29.3) into the Hamiltonian function (5) of the original canonical system (3), so that

$$h^0 = h^0(Y^0, t) \equiv h(F(Y^0, Y^*, t), t). \tag{67}$$

This implies, of course, that $h(F(Y^0, Y^*, t), t)$ is a function of Y^0 and t alone, that is, that

$$\operatorname{grad}_{Y^*} h(F(Y^0, Y^*, t), t) \equiv 0^{m-2l}. \tag{67 bis}$$

If a canonical parametrization (11) of (8) does not contain t, that is, if $F_t \equiv 0$, then the parametrization is completely canonical. In order to prove this, notice first that if $F_t \equiv 0^{2n}$, then, as seen from §3, the functions R and r of Y and t are independent of t. Hence it is sufficient to prove that r is independent of Y. In fact, r is then a constant and can, therefore, be omitted in (46). Now $\operatorname{grad}_Y r \equiv 0^m$ follows from (21.2) in view of $R_t \equiv 0^m$ and $F_t \equiv 0^{2n}$.

Remark. On comparing the fact thus proved with the considerations of §5, it is seen that every invariant system (8) which does not contain t explicitly admits a completely canonical parametrization (the Hamiltonian function (5) need not be independent of t).

As an application of completely canonical parametrizations, a generalization of the well known case of ignorable coordinates will now be considered.

If a coordinate $x_\lambda = q_\lambda$ (cf. (4), (4 bis)) does not occur in the Hamiltonian function (5) of (3), the corresponding impulse $x_{n+\lambda} = p_\lambda$ clearly is independent of t for *every* solution $X(t)$ of (3). Thus the case of an ignorable coordinate is the case of first integrals (cf. §10) of the form $x_\lambda = \text{const.}$ and is implied by the more general case where $x_\lambda = q_\lambda$ does occur in (5) but $x_{n+\lambda} = p_\lambda$ has a constant

† Cf. Levi-Civita [4].

value $c_{n+\lambda}$ for *some* solutions $X(t)$ of (3), so that $x_{n+\lambda}=c_{n+\lambda}$ is not a first integral but merely an invariant relation. This case of coordinates which are ignorable in a generalized sense is again a particular case of the one where a system (8) of $2n-m$ invariant relations is given as consisting on the one hand of $n-l$ relations of the form

$$x_{n+\lambda} = c_{n+\lambda}, \qquad \lambda = l+1, \cdots, n, \tag{68.1}$$

and on the other hand of a system of $m-n+l$ relations of the type

$$x_\mu = f_\mu(x_1, \cdots, x_{m-l}; x_{n+1}, \cdots, x_{2n}), \qquad \mu = m-l+1, \cdots, n, \tag{68.2}$$

where l denotes some integer satisfying (23 bis). The last assumption implies that

$$0 \leqq l \leqq n \quad \text{and} \quad m-l \leqq n \tag{68 bis}$$

and that all the x's occurring in (68.1) are impulses p in view of (4 bis). The case (68.1), (68.2) of (8) is of importance in certain applications† and can be parametrized in a completely canonical way as follows:

First, a parametrization (11) of the invariant system (8) represented by the m relations (68.1) and (68.2) may be obtained by choosing as the m components y of the parameter vector (10) those m components x_i of the $2n$-vector (4) for which either $1 \leqq i \leqq m-l$ or $n \leqq i \leqq n+l$. If one numerates the m components of this Y in a suitable manner, the parametrization (11) of the invariant system (68.1), (68.2) appears in the form (11), if one adjoins the relations

$$x_i = y_i,\ x_{n+i} = y_{l+i},\ x_j = y_{l+j},\ i = 1, \cdots, l\,;\, j = l+1, \cdots, m-l, \tag{68.3}$$

to (68.1), (68.2). Hence it is easily verified from (1) and (13) that the integer l occurring in (68.1), (68.2), (68.3) is the same as the one defined by (22), and that the matrix $J'GJ$ is, in view of (57), precisely the matrix (62). Since (62) is satisfied, the parametrization is canonical; that it is completely canonical follows from the fact that t does not occur in (68.1), (68.2), (68.3), that is, in (11).

Accordingly, the reduction of (3) to the separated canonical system (63) is, in the case of an invariant system of the form (68.1), (68.2), identical with the reduction of (3) in the classical case of first integrals represented by ignorable coordinates. In this particular case, the function (50 bis) is, in view of (68.3), independent of Y^*; hence $Y^* = Y^*(t)$ follows from (50) by quadrature, while the system (50) determining Y^* cannot be solved always by quadratures in the more general case (68.1), (68.2), (68.3).

† Cf. van Kampen and Wintner [3], pp. 155–156 and pp. 164–166.

8. **Canonical parametrization of holonomic invariant systems.** In this section the number l will be introduced as a given integer for which

$$0 \leqq l \leqq n, \tag{69}$$

but it will turn out that l is identical with the number defined by (22), while the number m defined by (7) will become

$$m = n + l. \tag{70}$$

The invariant system (8) of (3) will be called *holonomic* if it is independent of the n components of the $2n$-vector (4) which are represented in (4 bis) by the n-vector P; that is, if (8) is of the form

$$U(Q, t) = 0^{n-l}, \tag{71}$$

where

$$U = U(Q, t) \tag{72}$$

is an $(n-l)$-vector depending on the n-vector Q, defined by (4 bis) and possibly on t. It is clear from (4 bis) that in this holonomic case (7) does over into

$$n - l = \text{rank of } I, \tag{73}$$

where I denotes the Jacobian $(n-l, m)$-matrix

$$I = I(Q, t) = \text{grad}_Q\, U, \tag{74}$$

m being the number (70). According to (70), one can write the parameter vector (10) in the form

$$Y = \begin{pmatrix} \overline{Q} \\ \overline{P} \\ P^* \end{pmatrix}, \tag{75}$$

where $\overline{Q}$ and $\overline{P}$ are l-vectors and P^* is an $(n-l)$-vector. Furthermore, it is seen from (73) that (71) admits by means of an n-vector

$$A = A(\overline{Q}, t) \tag{76}$$

a parametrization of the form

$$Q = A(\overline{Q}, t), \tag{77}$$

where

$$l = \text{rank of } \text{grad}_{\overline{Q}}\, A. \tag{78}$$

This clearly implies the existence of (l, n)-matrices

$$H = H(\overline{Q}, t) \tag{79}$$

for which

$$H \operatorname{grad}_{\overline{Q}} A = E_l \tag{80}$$

(where E_l is the unit matrix) is an identity.

Now it will be shown that if an invariant system (8) of (3) is holonomic, then it admits a parametrization (11) which is canonical and of the form

$$\begin{pmatrix} Q \\ P \end{pmatrix} \equiv X = F(Y, t) = \begin{pmatrix} A \\ H'\overline{P} + I'P^* \end{pmatrix}, \tag{81}$$

where the m-vector Y is written in the form (75), the n-vector A is arbitrarily chosen such that (77) gives a parametrization of (71) satisfying (78), and H is any matrix satisfying (80); finally the matrix I defined by (74) is thought of as expressed by means of (77) as a function of $(\overline{Q}, t)$.

In order to prove this, notice first that substitution of (77) into (71) gives an identity in $(\overline{Q}, t)$. On differentiating this identity with respect to the components of the l-vector $\overline{Q}$, one obtains

$$I \operatorname{grad}_{\overline{Q}} A = 0^{n-l}_l \tag{82}$$

in view of (74). Now (80), (82), and (73) clearly imply that the (n, n)-matrix

$$\begin{pmatrix} H \\ I \end{pmatrix} \tag{83}$$

has the rank n; that is, it has a non-vanishing determinant. Hence it is clear from (75) that, since (77) is a parametrization of the restriction (71) for Q, (81) is a parametrization of the invariant system (8) represented by (71) and (4 bis). It remains to be shown that this parametrization of the holonomic invariant system is canonical.

First, the $(2n, n+l)$-matrix J defined by (13) is, according to (81) and (75),

$$J = \begin{pmatrix} \operatorname{grad}_{\overline{Q}} A & 0^n_l & 0^n_{n-l} \\ \operatorname{grad}_{\overline{Q}} (H'\overline{P} + I'P^*) & H' & I' \end{pmatrix}.$$

Hence it is seen from (1), (57) and (80), (82) that the matrix product $J'GJ$ satisfies the criterion (62) of a canonical parametrization if and only if the (l, l)-matrix

$$(\operatorname{grad}_{\overline{Q}} (H'\overline{P} + I'P^*))' \operatorname{grad}_{\overline{Q}} A - (\operatorname{grad}_{\overline{Q}} A)' \operatorname{grad}_{\overline{Q}} (H'\overline{P} + I'P^*)$$

is identically zero, that is, if and only if the (l, l)-matrix

$$D = (\operatorname{grad}_{\bar{Q}} A)' \operatorname{grad}_{\bar{Q}} (H'\overline{P} + I'P^*)$$

is symmetric. Now if one brings the factor $(\operatorname{grad}_{\bar{Q}} A)'$ of D beneath the grad sign of the second factor of D, one clearly obtains

$$D = \operatorname{grad}_{\bar{Q}} \{(\operatorname{grad}_{\bar{Q}} A)'(H'\overline{P} + I'P^*)\} + \text{correction term},$$

where the correction term, being an iterated gradient, is a symmetric matrix. Hence it is sufficient to show that

$$\operatorname{grad}_{\bar{Q}} \{(\operatorname{grad}_{\bar{Q}} A)'(H'\overline{P} + I'P^*)\}$$

is a symmetric matrix. Now this matrix is

$$\operatorname{grad}_{\bar{Q}} \{(H \operatorname{grad}_{\bar{Q}} A)'\overline{P} + (I \operatorname{grad}_{\bar{Q}} A)P^*\} = \operatorname{grad}_{\bar{Q}} \overline{P}$$

in view of (80) and (82); hence it is the zero matrix, since $\overline{P}$ is independent of $\overline{Q}$ (cf. (75)).

9. **Canonical parametrizations of semi-holonomic invariant systems.** The two types of invariant systems treated below under (i) and (ii) are important in some applications.† Each of these two invariant systems consists of two subsystems one of which is independent of the impulses while the other is linear in the impulses. One of the two subsystems, when considered without the other, need not be an invariant system. The number l will be introduced, in both cases (i), (ii), as a given integer satisfying (69); but it turns out that l is identical with the number l defined by (22), while (70) is replaced by

$$m = 2l. \tag{84}$$

The difference between the "semi-holonomic" cases (i), (ii) and the holonomic case of §8 depends, in the main, on the fact that in the semi-holonomic cases the $(n-1)$-vector P^* is missing in (75), so that the m-vector (10) is, in accordance with (84), of the form

$$Y = \begin{pmatrix} \overline{Q} \\ \overline{P} \end{pmatrix}, \tag{85}$$

where Q and P are l-vectors.

Case (i). Let

$$U = U(Q, t), \tag{86.1}$$

$$V = V(Q, t) \tag{86.2}$$

† Cf. van Kampen and Wintner [3], pp. 154–155, and the applications given in [3].

be a pair of functions such that U is an $(n-l)$-vector, V is an l-vector and, if I and H denote the Jacobian $(n-l, n)$- and (l, n)-matrices

$$I = I(Q, t) = \operatorname{grad}_Q U, \tag{87.1}$$

$$H = H(Q, t) = \operatorname{grad}_Q V, \tag{87.2}$$

the determinant of the (n, n)-matrix (83) is distinct from zero, which clearly implies that

$$n - l = \text{rank of } I, \tag{88.1}$$

$$l = \text{rank of } H. \tag{88.2}$$

Hence there exist $(n-l, n)$-matrices

$$M = M(Q, t) \tag{89}$$

such that

$$HM' = 0^l_{n-l} \tag{90}$$

is an identity in (Q, t) and

$$n - l = \text{rank of } M. \tag{91}$$

The non-vanishing of the determinant of the (n, n)-matrix (83) also implies that, if $\overline{Q}$ denotes the l-vector

$$\overline{Q} = V(Q, t), \tag{92}$$

there exists a unique n-vector (76) such that (78) is satisfied and (77) gives a parametrization (71) in such a way that (92) is an identity in $(\overline{Q}, t)$ in virtue of (77).

Now suppose that an invariant system (8) of (3) is given by the pair of conditions

$$U = 0^{n-l}, \tag{93.1}$$

$$MP = 0^{n-l}, \tag{93.2}$$

where U, M are given functions (86.1), (89) which satisfy, for some function (86.2), the conditions described above, and Q, P are the n-vectors defined by (4 bis). It will be shown that the invariant system (8) represented by (93.1), (93.2) admits a parametrization (11) which is canonical and of the form

$$\begin{pmatrix} Q \\ P \end{pmatrix} \equiv X = F(Y, t) \equiv \begin{pmatrix} A \\ H'\overline{P} \end{pmatrix}, \tag{94}$$

where the m-vector Y is written in the form (85) by means of two l-vectors

$\overline{Q}, \overline{P}, (m = 2l)$; the unique n-vector (76) defined before by means of (92) is such that (77) gives a parametrization of (93.1); the matrix H defined by (87.2) is thought of as expressed by means of (77) as function of (Q, t); and finally the inversion formula belonging to (93) is

$$Y \equiv \begin{pmatrix} \overline{Q} \\ \overline{P} \end{pmatrix} = \begin{pmatrix} V \\ (\mathrm{grad}_{\overline{Q}}\, A)'\overline{P} \end{pmatrix} \tag{95}$$

in virtue of the invariant system.

In fact, the first row of (95) and the first row of (94) are satisfied by the definitions (92) and (77) of $\overline{Q}$ and A. Furthermore, it is seen from the definition (87.2) of H that (80) is an identity in $(\overline{Q}, t)$ in virtue of (77). Now (80), (90), and (91) imply that the (n, n)-matrix

$$\begin{pmatrix} (\mathrm{grad}_{\overline{Q}}\, A)' \\ M \end{pmatrix} \tag{96}$$

has the rank n, that is, has a non-vanishing determinant. Hence the system

$$\begin{pmatrix} (\mathrm{grad}_{\overline{Q}}\, A)' \\ M \end{pmatrix} P = \begin{pmatrix} \overline{P} \\ 0^{n-l} \end{pmatrix} \tag{97}$$

of n linear equations has, for any l-vector $\overline{P}$, a unique solution P; and this P is

$$P = H'\overline{P} \tag{97 bis}$$

in view of (80) and (90). Hence, on considering the first row of (97) as the definition of $\overline{P}$, the second row of (94) follows from (93.2). This proves that the invariant system (8) represented by (93.1), (93.2) admits the parametrization (94) and satisfies (95). In order to prove that this parametrization is a canonical parametrization, that is, that the $(2l, 2l)$-matrix $J'GJ$, where J is defined by (13) and (94), satisfies the identity (62), one merely has to repeat, with obvious modifications, the verification carried out in detail at the end of §8.

Case (ii). Suppose that an invariant system (8) of (3) is given by a pair of conditions

$$U = 0^{n-l}, \tag{98.1}$$

$$MP = 0^{n-l}, \tag{98.2}$$

where Q, P are the n-vectors defined by (4 bis), and

$$U = U(Q, t), \tag{99.1}$$

$$M = M(Q, t), \tag{99.2}$$

(U is an $(n-l)$-vector, M is an $(n-l, n)$-matrix), and

$$n - l = \text{rank } IM', \tag{100}$$

where I denotes the $(n-l, n)$-matrix (74). The existence of a function (86.2), occurring in Case (i), is not assumed in the present Case (ii). Assumption (100) clearly implies (74) and (91), while (73) implies the existence of n-vectors (76) which satisfy (78) and are such that (77) gives a parametrization (98.1) in terms of an l-vector $\overline{Q}$.

It will be shown that the invariant system (8) represented by (98.1), (98.2) admits, for every choice of the function (76), a parametrization (11) which is canonical and of the form (94), (85), where the l-vector $\overline{P}$ is such that

$$\overline{P} = (\text{grad}_{\overline{Q}} A)'P, \tag{101}$$

while $H = H(\overline{Q}, t)$ is an (l, n)-matrix which is uniquely determined by the choice of (76) and satisfies (80) and (90).

In order to prove this, notice first that (82) is an identity in virtue of (77), the reason being the same as in §8. Now (82), (77), and (100) imply that the (n, n)-matrix (96) has the rank n, that is, it has a non-vanishing determinant. Hence there exists a unique (l, n)-matrix $H = H(\overline{Q}, t)$ which satisfies (80) and (90). Let (101) be considered as the definition of an l-vector $\overline{P}$. Since the determinant of (96) is distinct from zero, the linear system (97) has a unique solution P; and this P is given by (97 bis) in view of (80) and (90). This proves the second row of (94), while the first row of (94) is identical with (77). Finally, (98.2) and (101) are satisfied in view of (97). This proves that (98) is a parametrization (11) of the invariant system (8) represented by (98.1), (98.2). That this parametrization is a canonical parametrization, is proved in the same way as in Case (i).

10. **The case of first integrals.** The results of the preceding sections concern the general case of an invariant system (8) of (3). In what follows, it will be assumed that the relation

$$K(X, t) = C, \tag{102}$$

where C is a constant $(2n-m)$-vector, is an invariant system of (3) not only for a single choice of the constant vector C but for all choices of C in some $(2n-m)$-dimensional region of the C-space. This is the case if and only if each of the $2n-m$ scalar functions which constitute the components of the vector function $K = K(X, t)$ are first integrals of (3). Since a scalar function $g = g(X, t)$ represents a first integral of (3) if and only if $\dot{g} = 0$ is an identity in virtue of (3) alone, it is sufficient to prove that if (102) is an invariant system for every C, then $\dot{K} = 0^{2n-m}$ is an identity in virtue of (3) alone. Now if (102)

is an invariant system for every C, then $\dot{K}=0^{2n-m}$ is an identity in virtue of (3) and (102) together, and so, C being arbitrary, it is an identity in virtue of (3) alone.

In the particular case where the invariant system consists of first integrals, the integer $2l$ defined by (22) or (24) has a simple meaning, since it is connected with an integer occurring in the classical theory of reduction of canonical systems by means of first integrals. In fact, it turns out that if the $2n-m$ first integrals represented by (102), which are independent in view of (7), form a function group in the sense of Lie,* then this function group contains $n-l$, but not more, independent first integrals in involution. In view of the well known extension process of Poisson parentheses, the assumption that the $2n-m$ first integrals form a function group does not imply any loss of generality.

Since, on placing

$$2k = \text{rank of } NGN', \quad \text{where} \quad N = \operatorname{grad}_x K, \tag{103}$$

one can write (24), in view of (22), in the form

$$n - k = m - l, \tag{104}$$

the statement to be proved may be formulated as follows: If (102) consists of $2n-m$ first integrals of a canonical system (3) with n degrees of freedom, and if these first integrals are independent and form a function group, then

$$j = 2n - m - k, \tag{105.1}$$

where k is the integer defined by (103) and j denotes the maximum number of first integrals in involution which are contained in the function group.

First, NGN' is a skew-symmetric $(2n-m, 2n-m)$-matrix in which the j elements contained in the first j rows and first j columns vanish, if the first j components of the $(2n-m)$-vector $K(x, t)$ are independent first integrals in involution. Hence it is clear from (103) that

$$2k \leqq 2(2n - m - j), \quad \text{that is,} \quad j \leqq 2n - m - k. \tag{105.2}$$

In order to complete the proof of (105.1), it remains to be shown that

$$j \geqq 2n - m - k. \tag{105.3}$$

Now (105.3) is implied by a known existence theorem concerning functions in involution.†

* Cf., e.g., the presentation of Engel [1].

† The presentation of the local existence theorems by Engel [1] assumes, but actually does not use at all, that the systems are analytic.

Remark. There may be mentioned an essential difference between the case (8) of invariant systems and the case (102) of first integrals (in involution). In the case of first integrals one can solve (50) by mere quadratures. In other words, the separation of (19) into (48) and (50) reduces the degree of freedom from n to l in case of first integrals. This is a well known consequence of the Hamilton-Jacobi theory (Lie). On the other hand, the solution of (50) can require operations higher than quadratures in case of invariant systems.

Application to the problem of three bodies. In order to illustrate the use of (104) and (105.1), consider the non-planar case of the problem of three bodies. Let ξ_i, η_i, ζ_i, where $i=1, 2, 3$, be Cartesian coordinates of the mass μ_i in an inertial coordinate system, and let $\Xi_i, \mathrm{H}_i, \mathrm{Z}_i$ denote the conjugated impulses $\mu_i\dot{\xi}_i, \mu_i\dot{\eta}_i, \mu_i\dot{\zeta}_i$. Then $n=9$ and (3) admits, besides the energy integral, the nine classical first integrals represented by the functions

$$f_1 = \sum_{i=1}^{3} \{\mathrm{H}_i\zeta_i - \mathrm{Z}_i\eta_i\}, \quad f_2 = \sum_{i=1}^{3} \{\mathrm{Z}_i\xi_i - \Xi_i\zeta_i\}, \quad f_3 = \sum_{i=1}^{3} \{\Xi_i\eta_i - \mathrm{H}_i\xi_i\};$$

$$f_4 = \sum_{i=1}^{3} \Xi_i, \qquad f_5 = \sum_{i=1}^{3} \mathrm{H}_i, \qquad f_6 = \sum_{i=1}^{3} \mathrm{Z}_i;$$

$$f_7 = \sum_{i=1}^{3} \{\mu_i\xi_i - t\Xi_i\}, \quad f_8 = \sum_{i=1}^{3} \{\mu_i\eta_i - t\mathrm{H}_i\}, \quad f_9 = \sum_{i=1}^{3} \{\mu_i\zeta_i - t\mathrm{Z}_i\}.$$

Let the $(2n-m)$-vector K occurring in (105.1) be defined as the 9-vector with the components $f_1, \cdots, f_9$, so that $m=9$. Thus (104) and (105.1) go over into

$$l = k \quad \text{and} \quad j = 9 - k. \tag{106}$$

Now it will be easy to verify from (103) that $k=4$, so that $l=4$ and $j=5$ are, respectively, the degree of freedom of the separated system (63) and the maximum number of first integrals in involution contained in the function group generated by $f_1, \cdots, f_9$.

In order to prove that $k=4$, one has to show in view of (103) and (25.1) that the 9-rowed skew-symmetric matrix

$$\|(f_\rho, f_\sigma)\|; \qquad \rho, \sigma = 1, \cdots, 9,$$

of Poisson parentheses has the rank 8 if the values of the coordinates ξ_i, η_i, ζ_i and the impulses $\Xi_i, \mathrm{H}_i, \mathrm{Z}_i$ do not have exceptional values. Now it is easily verified from the above definition of $f_1, \cdots, f_9$ that

$$(f_1, f_2) = f_3, \qquad (f_2, f_4) = -f_6, \qquad (f_3, f_4) = f_5,$$

$$(f_2, f_7) = -f_9, \qquad (f_3, f_7) = f_8, \qquad (f_4, f_7) = \sum_{i=1}^{3} \mu_i,$$

and that these 6 identities remain valid under simultaneous cyclic permutations of the three triples (1, 2, 3), (4, 5, 6), (7, 8, 9); and finally that those $9^2-2\cdot 18=45$ Poisson parentheses which occur neither among these eighteen nor among their negatives $(f_\sigma, f_\rho)=-(f_\rho, f_\sigma)$, are identically zero. In other words, the skew-symmetric (9, 9)-matrix $\|(f_\rho, f_\sigma)\|$ may be written in the form

$$\|(f_\rho, f_\sigma)\| = \begin{pmatrix} F_1 & F_2 & F_3 \\ F_2 & 0 & \sum_{i=1}^{3} \mu_i E \\ F_3 & -\sum_{i=1}^{3} \mu_i E & 0 \end{pmatrix}, \tag{107}$$

where 0 and E denote the 3-rowed zero and unit matrices, respectively, and F_i is the three-rowed skew-symmetric matrix

$$F_i = \begin{pmatrix} 0 & f_{3i} & -f_{3i-1} \\ -f_{3i} & 0 & f_{3i-2} \\ f_{3i-1} & -f_{3i-2} & 0 \end{pmatrix}, \qquad i = 1, 2, 3. \tag{108}$$

It is easily seen from (108) and from the above representation of $f_1, \cdots, f_9$ that, if the eighteen coordinates and impulses have values which do not lie on an algebraical manifold of a dimension number less than eighteen, the 9-rowed skew-symmetric matrix (107) is of rank 8; that is, $k=4$.

If one considers the planar solutions of the problem of three bodies, one can assume that the three ζ_i and the three $\mathbf{Z}_i$ are identically zero, and there remain only the five integrals f_3, f_4, f_5, f_7, f_8. Thus $n=6$ and $2n-m=5$, so that (104) and (105.1) go over into

$$l = k+1, \; j = 5-k. \tag{109}$$

Since the 5-rowed skew-symmetric matrix which takes the place of the 9-rowed matrix $\|(f_\rho, f_\sigma)\|$ is found to be of rank 4 by the above method, it follows that $k=2$, and therefore, by (109), that $l=3$ and $j=3$.

11. **Forced paths.** So far it has been assumed that (8) is an invariant system of (3). In the present section it will only be assumed that (8) satisfies (7) and determines a non-empty manifold $\Lambda(t)$; that is, that (8) is compatible with itself, without necessarily being compatible with (3); so that (8) plays the role of a system of $2n-m$ constraints which modify the equations of motion (3). These constraints, in view of (4 bis), are not necessarily holonomic and are not given in the usual way, that is, not in terms of the coordinates, *velocities*, and time. It can be assumed, in view of (7), that the manifold (8) is represented by means of a parameter vector (10) in the form (11), (12).

Denote by

$$Y = Y(t) \tag{110}$$

any curve in the Y-space, with the time t as parameter, and let the curve in the X-space,

$$X = X(t), \tag{111}$$

which is defined by (110) and (14), be called the image curve of (110). Put, along this image path,

$$Z = Z(t) = G\dot{X} - \operatorname{grad}_X h, \tag{112}$$

so that the $2n$-vector $Z(t)$ is identically zero if and only if the image curve of (110) represents a solution of (3). The $(2n, m)$-matrix (13) is a function of t along the curve (110). Now if a curve (111) in the X-space is such that it is the image curve of a curve (110) for which

$$J'Z = 0^m \tag{113}$$

is an identity in t, then (111) will be termed a *forced path* belonging to the Hamiltonian function (5) and to the constraint system (8), since it is clear that whether a path (111) is a forced path or not is independent of the choice of the parametrization (11) of (8). It is seen from the identities (15), (36) and from the definitions (112) and (113) that a path (111) is a forced path belonging to (5) and (8) if and only if the differential condition

$$J'GJ\dot{Y} = -J'GF_t + \operatorname{grad}_Y h, \tag{114}$$

where h is defined by (35), has a solution (110) by means of which (111) is representable in the form (14). Since the skew-symmetric (m, m)-matrix $J'GJ$, defined in the $(m+1)$-dimensional (Y, t)-region under consideration, can have a rank (22) which is less than m, the system (114) cannot always be written in the normal form (19) of ordinary differential equations. In other words, one has general existence and uniqueness theorems for the initial problem of (114) only when accidentally $m = 2l$.

In what follows, only Hamiltonian functions (5) and constraints (8) will be considered for which the initial problem of (114) has not only an existence but also a uniqueness theorem in the region under consideration (to this end, the assumption $m = 2l$ is sufficient but not necessary). Then the dynamical meaning of forced paths is indicated by the following remark: *If one assumes* that the $2n$-vector (112) is a force, then the orthogonality condition (113) defining a forced path is nothing but d'Alembert's principle; which expresses the

fact that the forces Z, which modify the equations (3) of free motion in view of the constraint (8), do not work.

It is hard to verify the general validity of this interpretation. The interpretation is certainly valid in case (8) is not a constraint but an invariant system, since in this case (16) is valid and (114) follows from (36). Furthermore, the remark is valid in case the constraints are holonomic, since in this case (113) will be seen to be identical with the differential equations usually derived from d'Alembert's principle by means of Lagrangian multipliers. Whether or not the assumption (113) is dynamically correct in case of non-holonomic constraints, is a difficult, if not meaningless, question. In fact, non-holonomic constraints are usually assigned in terms of velocities and not of impulses, while X in (8) does contain impulses in view of (4 bis). Now, in case P actually occurs in (8), the passage from the impulse vector P to the velocity vector Q cannot be carried out before solving (113).

It remains to be verified that in the case of holonomic constraints which can contain t explicitly, the requirement (103) of a forced path leads to the same paths as d'Alembert's principle; that is, in the holonomic case the forced paths are identical with the paths described under the constraints which are required by the usual rules of dynamics. First, if (8) depends only on Q and t and not on P, the equations of motion belonging to the Hamiltonian function (3) and to the system (8) of $2n-m$ constraints are

$$\dot{Q} = \operatorname{grad}_P h, \qquad -\dot{P} = \operatorname{grad}_Q h + (\operatorname{grad}_Q K)'\lambda,$$

where λ is a $(2n-m)$-vector representing Lagrangian multipliers. This may be written, in view of (4 bis) and (1), in the form

$$G\dot{X} = \operatorname{grad}_X h + (\operatorname{grad}_X K)'\begin{pmatrix} 0^m \\ \lambda \end{pmatrix},$$

since $\operatorname{grad}_P K \equiv 0^n{}_{2n-m}$. On multiplying by J' and using the definition (112) of Z, one obtains

$$J'Z = J'(\operatorname{grad}_X K)'\begin{pmatrix} 0^m \\ \lambda \end{pmatrix}.$$

This relation is identical with the equation (113) of forced paths, since

$$(\operatorname{grad}_X K)J = 0^m{}_{2n-m}$$

is, according to (13), an identity in (Y, t) in virtue of (11).

Needless to say, the consideration of forced paths in case of holonomic constraints gives for d'Alembert's principle a formulation which is invariant under canonical transformations of the $2n$-dimensional phase-space.

References

1. F. Engel, *Die Liesche Theorie der partiellen Differentialgleichungen erster Ordnung*, edited by K. Faber, Leipzig, 1932.

2. E. R. van Kampen and A. Wintner, *On the canonical transformations of Hamiltonian systems*, American Journal of Mathematics, vol. 58 (1936), pp. 851–863.

3. E. R. van Kampen and A. Wintner, *On a symmetrical canonical reduction of the problem of three bodies*, American Journal of Mathematics, vol. 59 (1937), pp. 153–166.

4. T. Levi-Civita, *Sur la recherche des solutions particulières des systèmes différentiels et sur les mouvements stationaires*, Prace Matematyczno-Fizyczne, vol. 17 (1906), pp. 1–40.

5. T. Levi-Civita, *Sulla introduzione di vincoli olonomi nelle equazioni dinamiche di Hamilton* Atti del Reale Istituto Veneto, (8), vol. 18_1 (1916), pp. 387–395.

6. A. Wintner, *On the linear conservative dynamical systems*, Annali di Matematica, (4), vol. 13 (1934), pp. 105–112.

The Johns Hopkins University,
Baltimore, Md.

AN ISOMORPHISM BETWEEN LINEAR RECURRING SEQUENCES AND ALGEBRAIC RINGS*

BY
MARSHALL HALL

I. INTRODUCTION

1. The Thirteenth Century was in but its second year when Fibonacci (or Pisano) proposed a problem on the number of offspring of a pair of rabbits, whose solution led to the sequence of numbers now named after him. There is reason to believe that Fermat derived many of his arithmetic theorems from a knowledge of recurrences and, certainly, his celebrated Last Theorem may be stated as a problem on sequences. Lucas† was the first to make any extended researches on sequences, establishing a great many properties of certain second order sequences. Carmichael [1]‡ in 1920 made the first attack on sequences in general, and established their fundamental property of modular periodicity.

This paper undertakes a general survey of the modular properties of linear recurring sequences, beginning from the results of a paper by H. T. Engstrom and two by Morgan Ward.§ No problems on sequences are considered here which are not strictly modular, though questions on divisibility sequences‖ and their remarkable factorization properties are closely related.

The mechanism which the author uses for examining the properties of sequences is an isomorphism between the set of all sequences satisfying a fixed recurrence and a polynomial ring of operators. The isomorphism is not with the abstract ring but with a particular realization of it, and this is not especially surprising as a linear sequence is essentially an exponential function. This isomorphism may be derived from the theory of generating functions, and includes the fundamental identity used in Ward [11].

In Chapter II the isomorphism is set up and the basic properties of the ring are examined and their interpretation is given for the sequences. For example, the zero divisors of the ring correspond to the sequences which satisfy recurrences of lower order.

---

* Presented to the Society, September 3, 1936; received by the editors August 27, 1937.

† For the early bibliography see Dickson's *History of the Theory of Numbers*.

‡ Numbers in square brackets refer to the bibliography at the end of this paper.

§ Engstrom [3] and Ward [11], [12].

‖ Lucas [7], Lehmer [6], Hall [4], Ward [14].

Chapter III considers in some detail the periods modulus $p^j, (j=1,2,\cdots)$, of a fixed sequence and of all sequences. The structure of possible periods is thoroughly investigated and a method is given for its complete determination. The period patterns are shown to be dependent upon the ideal structure (or lattice) of the ring.

Chapter IV gives a similar dissection of the numeric patterns of sequences, and their relation to null sequences. Necessary and sufficient conditions that a sequence be null modulo m are found, and a very elegant criterion for "p-adically null" sequences is given.

Chapter V treats some questions on the distribution of modular residues in sequences. It is shown that the internal modular structure of sequences is intimately bound up with residual groups. A diophantine equation on distribution numbers which Ward found for third order sequences is shown to be one of a family of equations, and it is also shown that similar equations hold for sequences of any order.

Whereas Chapters III and IV would seem to exhaust the possibilities of the problems considered, Chapter V is only the introduction to some extremely recondite questions. The author will consider these further in another paper.

II. The isomorphism

2. Consider a linear recurrence of kth order

$$u_{n+k} = a_1 u_{n+k-1} + \cdots + a_k u_n \tag{2.1}$$

in which $a_1, a_2, \cdots, a_k$ are rational integers. Two operations on sequences (v_n) satisfying (2.1) may be defined:

I. *Sum*:

$$(v_n) + (w_n) = (v_n + w_n).$$

II. *Scalar product*:

$$t(v_n) = (tv_n), \qquad \text{any rational integer.}$$

In addition an operator x is defined:

III. $x(v_n) = (v_{n+1})$.

These three operations may be combined to yield a ring $R(x)$ of polynomial operators on the sequences, under which the sequences satisfying (2.1) are closed. This ring is easily shown to be associative, commutative, and distributive.

The characteristic polynomial of the recurrence (2.1), $f(x) = x^k - a_1 x^{k-1} - \cdots - a_k$, has the property that

$$f(x)(v_n) = (0) \tag{2.2}$$

for all sequences (v_n) satisfying (2.1). We may set up an isomorphism between the sequences satisfying (2.1) and the ring $R(x)$ by the correspondences:

Primary isomorphism:

$$1 \rightleftarrows (w_n), \qquad h(x) \rightleftarrows h(x)(w_n), \tag{2.3}$$

where (w_n) is the sequence defined by $w_0 = w_1 = \cdots = w_{k-2} = 0$, $w_{k-1} = 1$, which shall be called the unit sequence. It is a simple matter to verify that (2.3) actually defines a one-to-one correspondence between the sequences satisfying (2.1) and the ring $R(x)$ of polynomials modulo $f(x)$. In fact

$$(v_n) \rightleftarrows V(x) = v_0 x^{k-1} + (v_1 - a_1 v_0) x^{k-2} + \cdots + (v_{k-1} - \cdots - a_{k-1} v_0) \tag{2.4}$$

for any sequence (v_n). This polynomial $V(x)$ is the (unique) polynomial of degree less than k in the residue class modulo $f(x)$ corresponding to (v_n) by the isomorphism (2.3). We note in passing that the sequence (v_n) is integral if and only if the coefficients of the corresponding canonical polynomial $V(x)$ are integral.

It is sometimes convenient to use a secondary form of the isomorphism in considering single terms of the sequence (v_n).

Secondary isomorphism:

$$\begin{aligned} V(x) &\to v_0, \\ xV(x) &\to v_1, \\ &\cdots\cdots, \\ x^i V(x) &\to v_i. \end{aligned} \tag{2.5}$$

The secondary isomorphism is simply a correspondence between an element $h(x)$ of the ring $R(x)$ and the coefficient of x^{k-1} in the canonical representative of the residue class $h(x)$ modulo $f(x)$. The relations of (2.5) are justified by the definition III of the operator x and by (2.4).

3. If $f(x)$ is reducible, then $R(x)$ contains zero divisors, and if $f(x)$ has multiple factors, then $R(x)$ contains nilpotent elements. Hence the theory of fields may not be applied to $R(x)$ although a number of theorems on fields hold true. Again, $R(x)$ is not in general a maximal order, and the arithmetic of algebras is not directly applicable to the arithmetic of $R(x)$. But we may apply many theorems on algebraic number fields to $R(x)$ without reproving them because of the following theorem:

THEOREM 3.1. PRESERVATION OF IDENTITIES. *If $F(a_1, \cdots, a_k)$ is a polynomial in $a_1, \cdots, a_k$ with rational integral coefficients and if $F = 0$ whenever $a_1, \cdots, a_k$ are rational integers such that $f(x) = x^k - a_1 x^{k-1} - \cdots - a_k$ is irreducible, then $F \equiv 0$.*

If we take $f(x)$ arbitrary modulo p and irreducible modulo q, then $F=0$. Hence $F\equiv 0 \pmod{p}$ for $a_1, \cdots, a_k$ arbitrary modulo any p and $F\equiv 0$.

If we define the *norm* of $(v_n)\rightleftarrows V(x)$ by $N(v_n)=N[V(x)]$, then we have

$$N(v_n) = (-1)^{[k/2]} \begin{vmatrix} v_0, & v_1, \cdots, v_{k-1} \\ v_1, & v_2, \cdots, v_k \\ v_2, & v_3, \cdots, v_{k+1} \\ \cdots & \cdots \\ v_{k-1}, & v_k, \cdots, v_{2k-2} \end{vmatrix}. \tag{3.1}$$

Similarly for the *spur* (or *trace*)

$$S(v_n) = kv_{k-1} - (k-1)a_1v_{k-2} - \cdots - a_{k-1}v_0. \tag{3.2}$$

It is easily verified that the norm of the unit sequence is unity and that its trace is k.

A classical result of interest here is the following:

THEOREM OF KRONECKER. *A necessary and sufficient condition that* (v_n) *satisfy a recurrence of lower order than* (2.1) *is that* $N(v_n)=0$.

Hence, with respect to the recurrence of lowest order which (v_n) satisfies, $N(v_n)\neq 0$.

Another rational integer associated with a sequence (v_n) is its *container*. For some purposes this is more useful than the norm.

DEFINITION. *If* $(v_n)\rightleftarrows g(x)$ *then the container of* (v_n) *is the least positive rational integer which* $g(x)$ *divides.*

Since the rational integers which $g(x)$ divides form a modulus, they are all multiples of the container. There is a very close relationship between the container and the norm given by the following theorem:

THEOREM 3.2. *Precisely the same primes divide the container and the norm.*

If m is the container of $(v_n)\rightleftarrows g(x)$, then $h(x)g(x)=m$. Hence $N(h(x))\,N(g(x))=N(m)=m^k$ and so $N(v_n)\,|\,m^k$. But from the definition of the container $m\,|\,N(v_n)$. Hence every prime dividing m divides $N(v_n)$ and conversely.

III. PERIOD PATTERNS

4. If for a fixed τ and all $n\geqq n_0$

$$u_{n+\tau} \equiv u_n \pmod{m}, \tag{4.1}$$

then the least τ is said to be the period* of (u_n) modulo m and the least n_0

* The notation and terminology of this paper are in agreement with Ward [11] for the most part. I use "period" following Engstrom rather than Ward's "characteristic number."

the numeric. The principal results on periods and numerics are given in Carmichael [1], Engstrom [3], and Ward [11]. Here we shall not consider periods individually. We shall study the period patterns of sequences satisfying (2.1), where two sequences are said to have the same period pattern modulo p if their periods modulo p^j, $(j=1, 2, 3, \cdots)$, are the same.

Following Ward [11], we reduce the determination of periods to the solution for τ_i of congruences

$$(4.2) \qquad (x^{\tau_i} - 1)g(x) \equiv 0 \pmod{p^j, F_i(x)},$$

where $(u_n) \rightleftarrows g(x)$, and

$$(4.3) \qquad f(x) \equiv F_1(x)F_2(x) \cdots F_r(x) \pmod{p^j}, \quad F_i(x) \equiv h_i(x)^{e_i} \pmod{p},$$

the $h_i(x)$ being irreducible and distinct modulo p.*

THEOREM 4.1. *The partial period τ_i of $(u_n) \rightleftarrows g(x)$ modulo p^j is the exponent to which x belongs modulo the ideal A_j of all y satisfying $yg(x) \equiv 0 \pmod{p^j, F_i(x)}$.*

This theorem and its proof are obvious, but it seems desirable to emphasize from the start the relation of periods to ideal theory.

As we shall confine our attention to a single component $F_i(x)$, the subscript i will be omitted hereafter. Moreover we may whenever desirable restrict ourselves to *ordinary* sequences for which $g(x) \not\equiv 0 \pmod{p, F(x)}$. If (u_n) is not ordinary, then $g(x) \equiv p^s k(x) \pmod{p^j, F(x)}$, where $(v_n) \rightleftarrows k(x)$ is ordinary, and the period of (u_n) modulo p^j is the period of (v_n) modulo p^{j-s}.

If λ is the exponent to which x belongs modulo p, $h(x)$,† that is, the least solution of

$$(4.4) \qquad x^n \equiv 1 \pmod{p, h(x)},$$

and if r is determined by

$$(4.5) \qquad p^{r-1} < e \leqq p^r,$$

then $\nu = p^r\lambda$ is the principal period of (2.1) modulo p (Ward [11], Theorem 10.4). The principal period (by definition the period of the unit sequence) modulo p is a fortiori a period of any sequence modulo p. But no writer has considered the possibility that a sequence may have a period which is a proper divisor of ν. This can indeed happen and we shall later see that such periods appear in period patterns in a special role.

* This is the well known Schönemann decomposition. The $F_i(x)$ are unique and p-adically determinate.

† Information on the value of λ is given by a law of higher reciprocity due to F. K. Schmidt [9], pp. 165–166. For a particularly simple proof of this relation see O. Ore [8], pp. 272–273.

THEOREM 4.2. *The period of any ordinary sequence modulo p^i is of the form $p^i\lambda$.*

Since $p^{i-1}\nu = p^{r+i-1}\lambda$ is a general period modulo p^i (Engstrom [3], p. 217), it is sufficient to show that the period of an ordinary (u_n) is a multiple of λ. If the ideal A_i of Theorem 4.1 is the unit ideal, then $g(x) \equiv 0$ $(p^i, F(x))$ and (u_n) is not ordinary. Otherwise, since A_i contains the primary ideal $[p^i, F(x)]$, A_i must be contained in the prime ideal $[p, h(x)]$. Hence we must have

$$x^\tau \equiv 1 \text{ (modd } p, h(x)), \tag{4.6}$$

and τ is a multiple of λ.

Now ν may be the principal period not only modulo p, but also modulo p^σ, where the greatest σ will be called the *defect.** Again, for a particular sequence $(u_n) \rightleftarrows g(x)$, ν may be a period not only modulo p^σ but also modulo $p^{\sigma+\rho}$, where ρ is called the initial defect as it depends on the initial values. The values of ρ and σ determine all periods which are multiples of ν as given in the following:

THEOREM 4.3. *If $p^\sigma \neq 2$, and ν is a period of (u_n) modulo $p^{\rho+\sigma}$, then the period of (u_n) modulo p^i for $j > \rho+\sigma$ is $p^{i-\rho-\sigma}\nu$; when $p^\sigma = 2$, if 2ν is a period modulo exactly 2^s, then the period of (u_n) modulo 2^i is 2^{i-s+1} for $j > s$.*

For the method of proof see Ward [11], pp. 619–620. The statement of his Theorem 11.1 is inaccurate as it does not consider the possibility of periods which are proper divisors of ν. Note that ν is not said to be the period of (u_n) modulo $p^{\rho+\sigma}$ but merely one of the periods. Examples can be given in which the period of (u_n) modulo $p^{\rho+\sigma}$ is a proper divisor of ν.

Example. $u_{n+3} = 3u_{n+1} - u_n$ with $u_0 = 1$, $u_1 = 2$, $u_2 = 4$. Here $\nu = 6$ is a period of (u_n) modulo 3 but not modulo 9. The period of (u_n) modulo 3 is, however, 2.

5. From Theorem 4.2 the periods of (u_n) modulo p^i are among the numbers λ, $p\lambda$, $p^2\lambda$, $\cdots$. Let $p^{j(i)}$ be the highest power of p for which $p^i\lambda$ is a period of (u_n). Then the period of (u_n) modulo p^i is $p^i\lambda$, where i is the least number for which $j(i) \geqq j$. And from Theorem 4.3 the period pattern of (u_n) is completely determined given $j(0), \cdots, j(r)$. We may represent a period pattern graphically if, in the cartesian plane, we plot the points $(i, j(i))$ and connect successive points with straight lines. The resulting graph will be called the *period path* of (u_n).

In congruences (modd p^i, $F(x)$), the *partial container* plays the role of the container in $R(x)$. The partial container is zero or the least positive rational

* Ward [11], p. 622, gives some information which shows that the defect can be greater than unity only in certain rare cases. The determination of the defect is the generalization of the (unsolved) problem of the Fermat quotients $(a^{p-1}-1)/p$.

number which $g(x)$ divides modulo p^j, $F(x)$. Evidently the partial container must be p^{δ_1}, with $\delta_1 \leqq \delta$ if p^δ is the highest power of p dividing the container of $g(x)$. It may be shown without difficulty that the exponent δ_1 is the same whatever the value of j.

THEOREM 5.1. *If p^{δ_h} is the partial container of $x^{p^h\lambda}-1$ (modd p^j, $F(x)$) then $p^h\lambda$ is a period of some ordinary (v_n) modulo p^{δ_h}, but of no ordinary (u_n) modulo p^j for $j > \delta_h$.*

Let $h(x)(x^{p^h\lambda}-1) \equiv p^{\delta_h}$ (modd p^j, $F(x)$). If $h(x) \equiv 0$ $(p, F(x))$, then $h(x) \equiv pk(x)$ (modd p^j, $F(x)$) and $k(x)(x^{p^h\lambda}-1) \equiv p^{\delta_h-1}$ $(p^{j-1}, F(x))$ and p^{δ_h} is not the partial container of $x^{p^h\lambda}-1$ contrary to the assumption. Hence $h(x) \not\equiv 0$ $(p, F(x))$ and $(v_n) \rightleftarrows h(x)$ is ordinary. Moreover $p^h\lambda$ is a period of (v_n) modulo p^{δ_h} since $h(x)(x^{p^h\lambda}-1) \equiv 0$ $(p^{\delta_h}, F(x))$. Suppose that $p^h\lambda$ is a period of some $(u_n) \rightleftarrows g(x)$ modulo p^{δ_h+1}. Then $g(x)(x^{p^h\lambda}-1) \equiv 0$ $(p^{\delta_h+1}, F(x))$ and so $g(x)h(x)(x^{p^h\lambda}-1) \equiv 0$ $(p^{\delta_h+1}, F(x))$ or $g(x)p^{\delta_h} \equiv 0$ $(p^{\delta_h+1}, F(x))$, whence $g(x) \equiv 0$ $(p, F(x))$ and (u_n) is not ordinary.

If $f(x)$ has a root of unity among its roots, the container of some x^n-1 is zero and this theorem is without content in the form here given. In §6 it will be shown in what way the theory may be modified to cover this case.

The complete set of (ordinary) period paths modulus 3^j belonging to the recurrence

$$u_{n+4} = -u_{n+3} - 3u_{n+2} + 5u_{n+1} - u_n \tag{5.1}$$

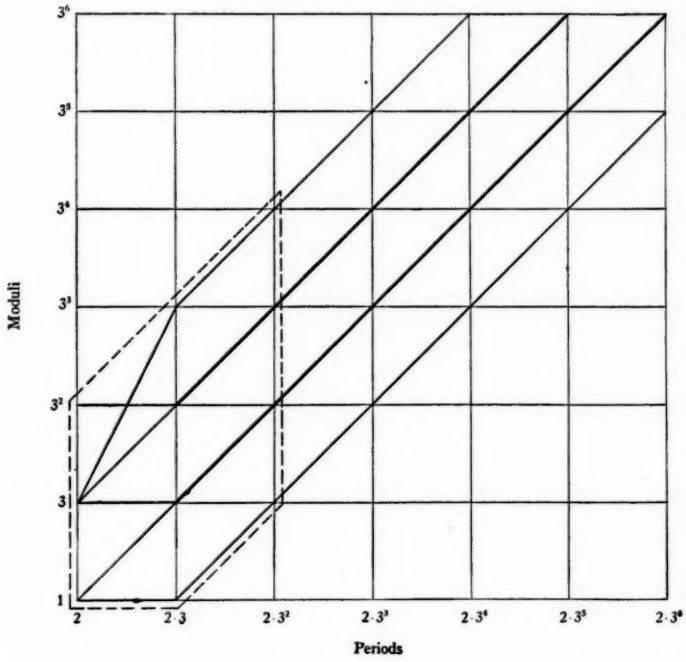
is given here.

There are six different period paths:

$(0, 0, 0, 1, \cdots) \rightleftarrows 1$ passing through	$(0, 0); (1, 0); (2, 1),$
$(0, 0, 1, 0, \cdots) \rightleftarrows x + 1$	$(0, 0); (1, 1); (2, 2),$
$(1, -1, -2, 11, \cdots) \rightleftarrows x^3 + 1$	$(0, 1); (1, 1); (2, 2),$
$(1, -1, -2, 8, \cdots) \rightleftarrows x^3 - 2$	$(0, 1); (1, 2); (2, 3),$
$(1, 2, -5, -1, \cdots) \rightleftarrows x^3 + 3x^2 - 5$	$(0, 1); (1, 3); (2, 4),$
$(1, -1, 1, 8, \cdots) \rightleftarrows x^3 + 3x + 1$	$(0, 2); (1, 2); (2, 3).$

Here, as $f(x) \equiv (x+1)^4$ (mod 3), we have $\lambda = 2$, $r = 2$, and $\nu = 18$. Theorem 4.3 shows that the period paths to the right of the line whose abscissa r corresponds to ν, are straight lines of inclination $\pi/4$.

The *period polygon* (dotted in the following figure) is defined as that polygon bounded on the left by the line $x = 0$, on the right by $x = r$, above by the line segments passing through the points (i, δ_i) in order, and below by the segments joining $(0, 0)$, $(r-1, 0)$, and (r, σ). Theorems 4.3 and 5.1 dispose of all period paths except those parts of ordinary period paths lying within the period polygon.



Principle of addition of paths. If the path of $g(x)$ passes through (i, j_1) and the path of $k(x)$ passes through (i, j_2), then the path of $g(x)+k(x)$ passes through (i, j'), where $j' = \min(j_1, j_2)$ when $i \neq j_2$ and $j' \geqq j_1$ when $j_1 = j_2$.

Proof. Suppose $j_1 < j_2$. Then $(x^{p^{2\lambda}}-1)g(x) \equiv 0 \ (p^{j_1}, F(x)) \not\equiv 0 \ (p^{j_1+1}, F(x))$ and $(x^{p^{2\lambda}}-1)k(x) \equiv 0 \ (p^{j_2}, F(x))$. Hence $(x^{p^{2\lambda}}-1)(g(x)+h(x)) \equiv 0 \ (p^{j_1}, F(x)) \equiv (x^{p^{2\lambda}}-1)g(x) \not\equiv 0 \ (p^{j_1+1}, F(x))$. For $j_1 = j_2$ the proof is obvious.

All $A(x)$ which satisfy $A(x)(x^{p^{i\lambda}}-1) \equiv 0 \ (p^i, F(x))$ form an ideal $A_{i,j}$. The path of an element belonging to $A_{i,j}$ will pass through or above (i, j). The path of (u_n) shows clearly to which $A_{i,j}$ $g(x)$ belongs and to which it does not belong.

An ideal is said to be ordinary if it contains an ordinary element.

Inclusion principle. If A is an ordinary ideal and is not included in the ideal B, then A contains an ordinary element which is not contained in B.

Proof. Choose an ordinary $h(x)$ from A. If $h(x)$ is in B, then we choose a $k(x)$ from A which is not in B. If $k(x)$ is ordinary, it satisfies our requirements. If $k(x)$ is not ordinary, then $h(x)+k(x)$ is ordinary and is in A but not in B.

Since the period of (u_n) modulo p^{i+1} is a multiple of the period modulo p^i, any period path must remain horizontal or go upward as we move to the right. We shall find further properties of the paths and give some existence theorems.

THEOREM 5.2. *There is an ordinary path passing through an arbitrary point of the period polygon.*

The lower boundary of the period polygon is the period path of the unit sequence. By Theorem 5.1 there is an ordinary path through an arbitrary point of the upper boundary. Note that as in the example of (5.1) the upper boundary is not necessarily a period path itself. If a point is s units above the lower boundary, then $(u_n) \rightleftarrows p^s$ passes through this point, and if we add this to the ordinary path passing through the point on the upper boundary immediately above it, we find an ordinary path passing through this point.

COROLLARY. $A_{i,j+1}$ *is properly contained in* $A_{i,j}$.

By actual calculation $A_{i,1} = [p, h(x)^{e-p^i}]$ for $i = 0, \cdots, r-1$. Also $p^{i-1}h(x)^{e-p^i}$ is in $A_{i,j}$ but not in $A_{i-1,j}$; hence $A_{i-1,j}$ is properly included in $A_{i,j}$. Using these simple facts, the principle of inclusion, and the principle of addition of paths, we may prove the existence of a variety of paths.

THEOREM 5.3. *There is an ordinary path through the following points of the period polygon*: (a) $(i-1, j)$, (i, j), *and* $(i+1, j)$ *provided that the last is not on the right boundary*; (b) $(i-1, j-1)$, (i, j), *and* $(i+1, j)$ *if all are interior points*; (c) $(i-1, j)$, (i, j), *and* $(i+1, j+1)$ *if all are interior points.* (d) *There is a path going up from every point on the lower boundary.* (e) *There is a horizontal path to the right from every point on the left or upper boundary.* (f) *There is an upward path from every point on the left boundary except possibly at the upper corner.* (g) *There is a path coming from below to every point on the right boundary, with one possible exception when* $p^\sigma = 2$, *and* $[p, h(x)] = [2, x+1]$.

The method of proof is the same in all cases, and only (g) will be proved here. Let the ordinary path (v_n) through (r, δ_r) pass through $(r-1, t)$. The unit sequence is ordinary and passes through $(r-1, 0)$ and (r, σ). If $(u_n) \rightleftarrows p^{j-\sigma}$, $(\sigma < j \leqq \delta_r)$, then $(u_n + v_n)$ is ordinary and passes through (r, j) and through $(r-1, j-\sigma)$ if $j-\sigma < t$, through $(r-1, t)$ if $j-\sigma > t$. Hence there is an exception only if $j-\sigma = t$ and (u_n+v_n) passes through $(r-1, j)$ and (r, j), that is, $(r-1, t+\sigma)$ and $(r, t+\sigma)$. If $\sigma > 1$, then $(w_n) \rightleftarrows p^{t+\sigma-1}$ passes through $(r-1, t+\sigma-1)$ and $(r, t+2\sigma-1)$; whence $(u_n+v_n+w_n)$ passes through $(r-1, t+\sigma-1)$ and $(r-1, t+\sigma)$. If $\sigma = 1$, it may happen that $(u_n+v_n+w_n)$ passes through $(r-1, t+\sigma-1)$ and (r, s) with $s > t+\sigma = t+1$. We now have an ordinary path $(y_n) \rightleftarrows y(x)$ through $(r-1, t+1)$ and $(r, t+1)$ and another

$(z_n) \rightleftarrows z(x)$ through $(r-1, t)$ and (r, s) with $s > t+1$ and no path to $(r, t+1)$ from below. (y_n+z_n) passes through $(r-1, t)$ and $(r, t+1)$ but may not be ordinary. If it is not ordinary, take $c \not\equiv 0$, 1 $(p, h(x))$; then the path of $y(x)+cz(x)$ is ordinary and passes through $(r-1, t)$ and $(r, t+1)$ as we wished. It is impossible to find such a c only if $[p, h(x)] = [2, x+1]$. In this case there may be an exception.

Example. If $F(x) = x^2+2x+5$, $p=2$, there is no ordinary path from below to (1, 3).

6. When $f(x)$ has one of its roots a root of unity, Theorem 5.1 may be without content, but all period paths may nevertheless be predicted by a finite process if we modify of the theory as in the following example:

$$u_{n+4} = -u_{n+3} + 3u_{n+2} - 4u_{n+1} + 2u_n. \tag{6.1}$$

Here $f(x) = x^4+x^3-3x^2+4x-2 = (x^2-x+1)(x^2+2x-2)$, and if we take $p=3$, then $\lambda = 2$ and $\nu = 18$.

$$\begin{aligned} (x^2 - 1)(5x^3 + 9x^2 - 6x + 17) &= -9, \\ (x^6 - 1)(x^2 + 2x - 2) &= 0, \\ (x^{18} - 1)(x^2 + 2x - 2) &= 0. \end{aligned} \tag{6.2}$$

Hence 2 may be a period modulo 1, 3, or 9, but 6 may be a period modulo 3^i for any i. Let $(u_n) \rightleftarrows g(x)$ be any sequence satisfying (6.1). Then we may write $g(x) = g_1(x)(x^2+2x-2)+g_2(x)$, where $g_2(x) = ax+b$. If (u_n) does not satisfy a recurrence of order lower than (6.1), then by the Theorem of Kronecker, $g_2(x)$ may not vanish since $N[g_1(x)(x^2-2x-2)] = 0$.

Let $g_2(x) = 3^i g_3(x)$, where $g_3(x) \not\equiv 0 \pmod 3$. Then $(x^6 - 1)g(x) = 3^i(x^6 - 1)g_3(x)$ or $(x^6 - 1)g(x) \equiv 3^i(x^2 - x + 1)(x^4 + x^3 - x - 1)g_3(x) \pmod{(x^2-x+1)(x^2+2x-2)}$. Hence we may reduce $(x^4+x^3-x-1)g_3(x)$ modulo x^2+2x-2, that is, to $(-11x+7)g_3(x)$, and as the partial container of $-11x+7 \pmod{3^i, x^2+2x-2}$ is 3, we find that $(x^6-1)g(x)$ is divisible either by 3^i or 3^{i+1}. Similarly the partial container of $(x^{18}-1)/(x^2-x+1) \pmod{3^i, x^2+2x-2}$ is 9. We may summarize these results by saying that 2 is a period modulo 1, 3, or 9; 6 is a period modulo 3^i or 3^{i+1}; and 18 is a period modulo 3^{i+1} or 3^{i+2}. For periods greater than 18, we may apply Theorem 4.3.

The methods used in this example are perfectly general and may be applied to any case in which the characteristic has roots of unity.

THEOREM 6.1. *If the greatest common divisor of $x^{p^h\lambda}-1$ and $F(x)$ is $r(x)$, then $p^h\lambda$ is a period of $(u_n) \rightleftarrows g(x)$ modulo p^t, $(i \leq t \leq i+\delta_h)$, where $g(x) \equiv p^i g_1(x) \pmod{p^i, F(x)/r(x)}$ and p^{δ_h} is the partial container of $(x^{p^h\lambda}-1)/r(x) \pmod{p^i, F(x)/r(x)}$.*

7. If $p^\sigma = 2$, we have not only the possibility of an exception to Theorem 5.3, but Theorem 4.3 is applicable only to periods greater than 2ν. Closer study of this case yields some interesting results.

In this case we will have

$$x^\nu - 1 \equiv 2M(x) \text{ (modd } 2^i, F(x)), \tag{7.1}$$

where $M(x) \not\equiv 0$ $(2, F(x))$. By direct calculation we obtain

$$x^{2\nu} - 1 \equiv 4M(x)[M(x) + 1] \quad (2^i, F(x)). \tag{7.2}$$

If $g(x)M(x)[M(x)+1] \equiv 2^t S(x)$ $(2^i, F(x))$, then 2ν is a period of $(u_n) \rightleftarrows g(x)$ modulo 2^{t+2}. But of the two quantities $M(x)$, $M(x)+1$, one or the other must be relatively prime to the modulus. Suppose $M(x)+1$ is relatively prime to the modulus. Then $g(x)M(x) \equiv 2^t S(x)/(M(x)+1) \equiv 2^t S_1(x)$ $(2^i, F(x))$, and from (7.1) $g(x)(x^\nu - 1) \equiv 2^{t+1} S_1(x)$ $(2^i, F(x))$, whence ν is a period of (u_n) modulo 2^{t+1}. In this case the period paths to the right of $x = r$ (corresponding to period ν) are straight lines of inclination $\pi/4$, that is, Theorem 4.3 may be applied for all periods greater than ν, as in the case of paths when $p^\sigma \neq 2$.

Suppose, on the other hand, that $M(x)+1$ is not relatively prime to the modulus. In this case $M(x)$ is relatively prime and its partial container is 1. Hence from (7.1) every ordinary period path goes through $(r, 1)$. Let $M(x)+1 \equiv 2^\kappa T(x)$, $T(x) \not\equiv 0$ $(2, F(x))$, and let 2^λ be the partial container of $T(x)$ (modd $2^i, F(x)$). Then it is easily shown by previous methods that there are ordinary paths passing through $(r+1, i)$ with $2+\kappa \leqq i \leqq 2+\kappa+\lambda$. An example of this is given by $F(x) = x^4 + 3x^2 + 4x - 3$. Here $\nu = 6$,

$$x^6 - 1 = 2(-2x^3 + 6x^2 + 6x - 5) = 2M(x). \tag{7.3}$$

Here $M(x)+1 = 2T(x)$, where the partial container of $T(x)$ is 2. The graph of the period paths has a curious "elbow" in it.

IV. Numerics and null sequences

8. The numeric of a sequence (u_n) modulo m is the least integer $n(m)$ for which $u_{n+\tau} \equiv u_n \pmod m$, $n \geqq n(m)$, where τ is the period of (u_n) modulo m. The null index (if it exists) of a sequence (u_n) modulo m is the least integer $n_1(m)$ for which $u_n \equiv 0 \pmod m$, $n \geqq n_1(m)$ and is of course also the numeric. Theorem 4.1 of Ward [11] shows that the existence of a sequence with a prescribed numeric implies the existence of another sequence with the same null index. This theorem may be extended to include a prescribed set of numerics for a finite number of moduli.

Let $f(x) \equiv F(x)G(x) \pmod{p^i}$, where $F(x) \equiv x^e \pmod p$ and $G(x) \not\equiv 0$ (p, x). This is the Schönemann decomposition of $f(x)$ with the factors $F_2(x) \cdots F_r(x)$

combined into the single term $G(x)$. The congruence for the numeric of (u_n) modulo p^j is

$$x^n g(x) \equiv 0 \text{ (modd } p^j, F(x)). \tag{8.1}$$

THEOREM 8.1. *A necessary and sufficient condition that* $(u_n) \rightleftarrows g(x)$ *be null modulo* p^j *is that* $g(x) \equiv 0$ (modd $p^j, G(x)$).

This theorem is given in Ward [11], pp. 613–614.

Let us call a sequence *p-adically null* if it is null modulo p^j for all j.

THEOREM 8.2. *If* p^δ *is the highest power of* p *dividing the container of* $(u_n) \rightleftarrows g(x)$, *then* (u_n) *cannot be null modulo* p^j *for* $j > \delta$ *unless* $G(x) = 1$, *in which case* (u_n) *is p-adically null.*

Let $g(x)h(x) \equiv p^\delta r \pmod{f(x)}$ where $r \not\equiv 0 \pmod{p}$. Then a fortiori $g(x)h(x) \equiv p^\delta r$ (modd $p^j, G(x)$) and hence $g(x) \not\equiv 0$ $(p^j, G(x))$ for $j > \delta$ unless $G(x) = 1$, in which case $g(x) \equiv 0$ (modd $p^j, G(x)$) for all j.

A less precise but more striking way of stating the same theorem is the following:

THEOREM 8.3. *If* (2.1) *is the recurrence of lowest order which* (u_n) *satisfies, then* (u_n) *is p-adically null if and only if p divides all of* $a_1, \cdots, a_k$.

It is possible to graph numeric paths of sequences in a way similar to that in which period paths were graphed. If n is a numeric of (u_n) modulo p^j but not modulo p^{j+1}, we plot the point (n, j). The broken line joining these points for $n = 0, 1, 2, \cdots$ is the numeric path of (u_n). There is an analogue to Theorem 5.1 for numeric paths, but not to Theorem 4.3. We note that a sequence is ordinary $[g(x) \not\equiv 0 \ (p, F(x))]$ if it is *not* purely periodic modulo p.

THEOREM 8.4. *The numeric path of the unit sequence is the lower boundary of all numeric paths, and the segments joining the points* (i, δ_i), *where* p^{δ_i} *is the partial container of* x^i *modulo* $p^j, F(x)$, *form the upper boundary of ordinary numeric paths.*

The proof is straightforward and parallels that of Theorem 5.1. The portion of the plane included between the lower and upper boundaries of ordinary numeric paths will be called the numeric sector.

THEOREM 8.5. *Let* $F(x) = x^e + b_1 x^{e-1} + \cdots + b_e$, *and let* p^{α_i} *be the highest power of* p *dividing* b_i. *If we define* $\alpha = \alpha_i/i = \min(\alpha_1, \alpha_2/2, \cdots, \alpha_e/e)$, *then a point on the lower boundary of the numeric sector is either on or within a distance* $e\alpha_i$ *below the ray through the origin whose slope is* α. *If* $\beta = \beta_j/j = \min(\alpha_{e-1}, (\alpha_{e-2}+\alpha_e)/2, \cdots, (e-1)\alpha_e/e)$, *then any point on the upper boundary is either on or within a distance* $e\beta_j$ *above the ray through the origin whose slope is* $\alpha_e - \beta$.

We may define a commutative algebra $\mathfrak{A}(x)$ over the field of all algebraic numbers by taking $1, x, \cdots, x^{e-1}$ as basis elements and putting $F(x)=0$. The element $y=p^{-\gamma}x$ is integral in the sense of Deuring [2] (p. 68) if γ is a rational number less than or equal to α, but not for any γ greater than α. Hence $z=p^{-\alpha_i}x^i=y^i$ satisfies an equation

$$z^s + c_1 z^{s-1} + \cdots + c_s = 0 \tag{8.2}$$

of degree not greater than e, where the c's are rational since z is a rational function of x, and are integral since z is an integral element of $\mathfrak{A}(x)$. Not all c's are divisible by p, for if they were, then $p^{-1/s}z$ would be an integral element of $\mathfrak{A}(x)$ and so also would be some $p^{-\gamma}x$ with $\gamma > \alpha$.

If (u_n) is the unit sequence satisfying a recurrence of characteristic $F(x)$, then by the secondary isomorphism (2.5)

$$x^{ni+\rho} \rightarrow u_{ni+\rho}. \tag{8.3}$$

Hence since $p^{\alpha_i}z=x^i$, we may write

$$p^{n\alpha_i}z^n x^\rho \rightarrow u_{ni+\rho}. \tag{8.4}$$

The elements $(x^\rho z^n)$, $n=0, 1, 2, \cdots$ and ρ fixed as one of $0, 1, \cdots, i-1$, satisfy the recurrence

$$v_{n+s} = -c_1 v_{n+s-1} - \cdots - c_s v_n. \tag{8.5}$$

Hence the leading coefficients in their minimal residues modulo $F(x)$ must also satisfy the same recurrence. But from (8.4) these are the sequence $(u_{ni+\rho}/p^{n\alpha_i})$, $(n=0, 1, \cdots)$. Now the coefficients of (8.5) are rational integers and hence the denominators of a sequence satisfying this recurrence cannot exceed the denominators of the initial values. Hence the denominator of any one of the sequences $(u_{ni+\rho}/p^{n\alpha_i})$, $(\rho=0, 1, \cdots, i-1)$, cannot exceed $p^{(e-1)\alpha_i}$ and a fortiori $p^{(e-1)\alpha_i}$. Hence $u_{ni+\rho}=p^{n\alpha_i}v_n=p^{n\alpha_i-\sigma}v_n^*$, where v_n^* is integral and $\sigma \leqq (e-1)\alpha_i$. As this relation holds for $\rho=0, \cdots, i-1$, we have $x^{ni+\rho} \equiv 0 \ (p^{n\alpha_i-\sigma})$. Hence the lower boundary passes through or above the points $(ni+\rho, n\alpha_i-\sigma)$, and as $\rho<i$, $\sigma<(e-1)\alpha_i$, these points are not more than $e\alpha_i$ below the ray through the origin with slope $\alpha=\alpha_i/i$. On the other hand, no point of the lower boundary can lie above this ray, for if $x^n \equiv 0 \ (p^j)$ with $j > n\alpha$, then $p^{-j}x^n$ is an integral element of $\mathfrak{A}(x)$ and so also is $(p^{-j}x^n)^{1/n}$ or $p^{-j/n}x$. But $p^{-\gamma}x$ cannot be an integral element of $\mathfrak{A}(x)$ with $\gamma > \alpha$.

From Theorem 8.4 to determine the upper boundary it is necessary only to find the partial container of x^n (modd $p^i, F(x)$). We have

$$x(x^{e-1} + b_1 x^{e-2} + \cdots + b_{e-1}) = -b_e = p^{\alpha_e} b \tag{8.6}$$

where $p \nmid b$. We may write this

$$xw = p^{\alpha_e} b. \tag{8.7}$$

Hence if $w^n = p^{\epsilon}\overline{w}$ with $p \nmid w$, then $p^{n\alpha_e - \epsilon}$ is the partial container of x^n. It remains to find the greatest power of p dividing w^n. But $w = b_e/x$ satisfies the rank equation

$$w^e + b_{e-1}w^{e-1} + \cdots + b_e^{e-2}b_1 w + b_e^{e-1} = 0. \tag{8.8}$$

Reasoning for w^n from this equation as for x^n from $F(x) = 0$, we find that $w^n \equiv 0 \pmod{p^{n\beta - \sigma}}$, where $0 \leqq \sigma \leqq e\beta_i$. Hence the partial container of x^n is $p^{n\alpha - n\beta + \sigma}$. This shows that a point on the upper boundary of the numeric sector is either on or within $e\beta_i$ units above the ray through the origin with slope $\alpha_e - \beta$. This completes the proof of the theorem, which incidentally justifies the use of the term "sector" for the portion of the plane containing ordinary numeric paths.

9. We now calculate the numeric paths.

THEOREM 9.1. *If p^δ is the partial container of $(u_n) \rightleftarrows g(x)$ modulo p^i, $F(x)$, then the numeric of (u_n) modulo p^i is at least as great as the numeric of the unit sequence modulo $p^{i-\delta}$.*

COROLLARY. *The numeric path of (u_n) remains within a distance δ above the lower boundary of the numeric sector.*

Proof. If n is the numeric of (u_n) modulo p^i, then by (8.1)

$$x^n g(x) \equiv 0 \ (\text{modd } p^i, F(x)), \tag{9.1}$$

and if $g(x)h(x) \equiv p^\delta \ (\text{modd } p^i, F(x))$, then multiplying (9.1) by $h(x)$ we obtain

$$x^n p^\delta \equiv 0 \ (\text{modd } p^i, F(x)), \tag{9.2}$$

whence $x^n \equiv 0 \ (\text{modd } p^{i-\delta}, F(x))$ and n must be at least as great as the numeric of the unit sequence modulo $p^{i-\delta}$, which is the least solution of this congruence.

In consequence of (9.1) we consider the coefficients of

$$[x^{ni+\rho}g(x)]/p^{n\alpha i}, \qquad \rho = 0, \cdots, i-1; n = 0, 1, 2, \cdots, \tag{9.3}$$

which for ρ fixed, by a previous argument, satisfy recurrence (8.5). As before, the denominators cannot exceed $p^{e\alpha i}$. The coefficients of

$$x^{ni+\rho}/p^{n\alpha i} \tag{9.4}$$

cannot for any single value of n all be divisible by p, since then the path of the unit sequence would come above the ray through the origin with slope α. Similarly, by the corollary to Theorem 9.1 the coefficients of (9.3) cannot for

a single value of n all be divisible by $p^{\delta+1}$. Hence to determine the numeric path of (u_n), we need only know the values of the coefficients of (9.3) modulo p^{δ}. But this is equivalent to knowing the values of the coefficients of $p^{e\alpha_i}[x^{n_i+\rho}g(x)]/p^{n\alpha_i}$ modulo $p^{e\alpha_i+\delta}$ and these are sequences of integers satisfying (8.5). We are thus reduced to a finite problem, since the numeric and period of a sequence modulo $p^{e\alpha_i+\delta}$ are finite. But in general δ is not bounded for the set of all ordinary sequences satisfying the recurrence.

10. The following properties of period paths may easily be verified to hold for numeric paths:

(a) As we move to the right, the path never moves downward.

(b) Every non-ordinary numeric path is an upward translation of an ordinary path.

(c) The sum path of two paths passes through the lower of the two and through or above points of intersection.

(d) The elements corresponding to paths passing through or above a point (n, j) form an ideal $B_{n,j}$.

Using these properties and Theorem 8.4, we easily deduce the following theorem:

THEOREM 10.1. *Through an arbitrary point of the numeric sector there is an ordinary path.*

There is no analogue to Theorem 4.3 for numeric paths, but there is an important relation of another sort.

THEOREM 10.2. *If any finite section at the beginning of a numeric path is cut off, the remainder is a translation of an entire ordinary numeric path.*

Suppose the numeric path of $(u_n) \rightleftarrows g(x)$ passes through the point (n, j). We have

$$x^n g(x) = p^j h(x), \tag{10.1}$$

where $h(x) \not\equiv 0 \pmod{p}$. The section of the numeric path of (u_n) beyond (n, j) is the numeric path of $(v_n) \rightleftarrows h(x)$. For if

$$x^{n+m} g(x) = p^{i+j} t(x), \tag{10.2}$$

where $t(x) \not\equiv 0 \ (p)$, then $x^m[x^n g(x)] = x^m[p^j h(x)] = p^{i+j} t(x)$ or

$$x^m h(x) = p^i t(x), \tag{10.3}$$

and conversely.

One consequence of this theorem is that no small section of a numeric path can have peculiarities not exhibited by a path in the neighborhood of the origin. For example, no path may at any point rise more rapidly than the first

segments of the upper boundary or more slowly than the first segments of the lower boundary. The existence of a variety of path structures may be deduced from these two fundamental theorems by adding appropriate paths.

11. In one particular instance we may by a finite process predict all numeric paths. By Theorem 8.5 and the corollary to Theorem 9.1 there will certainly be infinitely many numeric paths when the upper ray is distinct from the lower ray. But when these coalesce there are only a finite number of paths. This happens only when $\alpha+\beta=\alpha_e$. But $\alpha\leqq\alpha_e/e$ and $\beta\leqq(e-1)\alpha_e/e$. Hence we must have $\alpha=\alpha_e/e$ and $\beta=(e-1)\alpha_e/e$. We may take $z=x^e/p^{\alpha_e}$ and we find that $N(z)=(b_e)^e/p^{e\alpha_e}\not\equiv 0\ (p)$. Hence in the recurrence (8.5) $c_s\not\equiv 0\ (p)$ and the (v_n) sequences are purely periodic modulo p^i. Consequently the lower boundary of the numeric sector must touch the ray infinitely often. If it passes through a point (n, j) on the ray,

$$x^n \equiv p^j h(x)\ (F(x)), \tag{11.1}$$

where $h(x)\not\equiv 0\ (p, F(x))$ and $j/n=\alpha$. If $h(x)\equiv 0\ (p, x)$, then $h(x)^e\equiv 0\ (p, x^e)\equiv 0\ (p, F(x))$. Then $x^{ne}\equiv p^{je}h(x)^e\equiv p^{je+1}s(x)\ (F(x))$. But $(je+1)/ne>\alpha$ and this would mean that the lower path goes above the ray. Hence $h(x)\not\equiv 0\ (p, x)$ and the partial container of x^n must be p^j, and the upper boundary must also pass through (n, j). Similarly the lower boundary must pass through every point on the ray through which the upper boundary passes. The boundaries form a series of sausage-like loops about the ray.

Let (r, i) be a point on the numeric path of $(u_n)\rightleftarrows g(x)$. Then

$$x^r g(x) \equiv p^i t(x)\ (F(x)), \tag{11.2}$$

where $t(x)\not\equiv 0\ (p, F(x))$. Multiplying (11.1) by $x^r g(x)$ we obtain

$$x^{n+r}g(x) \equiv p^{j+i}t(x)h(x)\ (F(x)), \tag{11.3}$$

and here $t(x)h(x)\not\equiv 0\ (p, F(x))$ since $h(x)$ is relatively prime to the modulus and $t(x)\not\equiv 0\ (p, F(x))$. Hence $(n+r, j+i)$ is on the numeric path of (u_n) and similarly $(2n+r, 2j+i)$ and so on. Consequently the entire path structure is given by that in the first loop.

12. In the study of both period and numeric patterns, it has been shown that the pattern of $(u_n)\rightleftarrows g(x)$ depends on which ideals of the $A_{i,j}$ (or $B_{n,j}$) contain $g(x)$ and which do not. These are a subset of the primary ideals contained in a single prime ideal $[p, h(x)]$. Hence the following equivalence criterion is of interest:

THEOREM 12.1. *A necessary and sufficient condition that elements y and z of $R(x)$ belong to the same primary ideals contained in $[p, h(x)]$ is that they have the same partial container p^s (modd $p^i, F(x)$) and that*

$$z \equiv yu \ (p^\delta, F(x)), \qquad y \equiv zv \ (p^\delta, F(x)). \tag{12.1}$$

Necessity. Let the partial containers of y and z be p^{δ_1} and p^{δ_2}, respectively. If $\delta_1 \neq \delta_2$, suppose $\delta_1 < \delta_2$. Then $[p^{\delta_2}, z, F(x)]$ does not contain y. For if it does, then

$$y \equiv bz \ (p^{\delta_2}, F(x)). \tag{12.2}$$

There exist $w \not\equiv 0 \ (p, F)$ with $yw \equiv p^{\delta_1} \ (p^i, F)$ and $s \not\equiv 0 \ (p, F)$ with $zs \equiv p^{\delta_2} \ (p^i, F)$. Hence $swy \equiv bswz \ (p^{\delta_2}, F)$ and so

$$p^{\delta_1} s \equiv 0 \ (p^{\delta_2}, F); \qquad s \equiv 0 \ (p^{\delta_1 - \delta_2}, F) \tag{12.3}$$

contradicting $s \not\equiv 0 \ (p, F)$. Consequently $\delta_1 = \delta_2 = \delta$. Now the supposition that z is in $[p^\delta, y, F(x)]$ and that y is in $[p^\delta, z, F(x)]$ leads to the congruences (12.1).

Sufficiency. Any primary ideal A contained in $[p, h(x)]$ must contain some ideal $[p^i, F(x)]$ with a sufficiently large j. From (12.1) and $yw \equiv p^\delta \ (p^i, F(x))$

$$\begin{aligned} z &\equiv yu + p^\delta r \ (p^i, F(x)) \\ &\equiv yu + ywr \\ &\equiv y(u + wr) \\ &\equiv yu^*. \end{aligned} \tag{12.4}$$

Similarly $y \equiv zv^* \ (p^i, F(x))$. A fortiori

$$z \equiv yu^* \pmod{A}, \qquad y \equiv zv^* \pmod{A}; \tag{12.5}$$

hence $y \equiv 0 \pmod{A}$ implies $z \equiv 0 \pmod{A}$, and conversely.

V. Cycles and residual groups

13. The residues of a sequence (u_n) modulo m consist of a numeric section, $u_0, u_1, \cdots, u_{n_0-1}$, where n_0 is the numeric modulo m, and a periodic section $u_{n_0}, \cdots$. The periodic section consists of infinitely many repetitions of the *cycle* $u_{n_0}, \cdots, u_{n_0+\tau-1}$, where τ is the period modulo m. Any set of τ consecutive terms from the periodic section will be called a cycle. Each cycle is a cyclic permutation of any other. For most purposes it is convenient not to consider these cycles as distinct, and in this sense a sequence (u_n) has but one cycle modulo m. Whenever it is important to distinguish these, we shall call any particular one of them an *ordered cycle*. For the rest of the terminology used in this section, see Ward [12].

Here we shall restrict ourselves to the case in which the modulus is a prime p. If $p \nmid a_{k-r}$, $p \mid a_{k-r-1}, \cdots, p \mid a_k$, then the numeric section is included in $u_0, \cdots, u_{r-1}$ and these terms, being initial values, are arbitrary. Hence, without loss of generality, (u_n) may be supposed purely periodic, as the periodic section is independent of the numeric section. In addition, as the Theo-

rem of Kronecker holds true modulo p, with respect to the recurrence of lowest order which (u_n) satisfies modulo p, p is not a divisor of $N(u_n)$. Hence (Ward [11], p. 604) the period of (u_n) modulo p is the principal period.

In this section we shall always use the secondary isomorphism

$$x^n g(x) \rightarrow u_n, \tag{13.1}$$

and a cycle will correspond to the elements $x^n g(x)$, $(n=0, \cdots, \tau-1)$.

THEOREM 13.1. *The elements corresponding to a cycle of (u_n) form a coset of the cyclic group $\{x\}$ in the group of residues relatively prime to the modulus $p, f(x)$.*

Proof. The elements relatively prime to the modulus p, $f(x)$ form an abelian multiplicative group. As we are assuming (u_n) to be purely periodic, $p \nmid a_k$ and the element x is in this group, and so the cyclic group $\{x\}$ is also in this group. *As* $p \nmid N(u_n)$, $g(x)$ is relatively prime to the modulus. $x^\tau \equiv 1$ (modd $p, f(x)$) is the congruence for the (principal) period modulo p. Hence $\{x\}$ consists of the elements $1, x, \cdots, x^{\tau-1}$, and the elements $g(x)$, $xg(x), \cdots, x^{\tau-1}g(x)$ form a coset of this subgroup.

As a finite abelian group, the residues relatively prime to p, $f(x)$ are characterized completely by the generators and order invariants of the group. By a slight modification of the methods used in T. Takenouchi [10] these may be found explicitly. A knowledge of the generators and invariants will be presupposed here. Except when $f(x)$ has factors modulo p of high multiplicity they are very simple.

THEOREM 13.2. *If $y^n g(x) \rightarrow v_n$ by the secondary isomorphism, then (v_n) satisfies a recurrence whose characteristic is the minimal polynomial of y.*

For let

$$F(y) = y^s - b_1 y^{s-1} - \cdots - b_s = 0. \tag{13.2}$$

Certainly y, as a polynomial in x, satisfies an equation of degree k, but it may possibly satisfy an equation of lower degree. The elements $y^n g(x)$, $(n=0, 1, \cdots)$, satisfy

$$v_{n+s} = b_1 v_{n+s-1} + \cdots + b_s v_n \tag{13.3}$$

and hence their leading coefficients (v_n) must also satisfy it.

In studying cycles and the distribution of residues therein, we shall use not only the group G of residues prime to $p, f(x)$ (Theorem 13.1) and auxiliary recurrences (Theorem 13.2) but also theorems on the existence of sequences with zeros in prescribed positions, of which the two following are typical.

THEOREM 13.3. *There exists a sequence (u_n) satisfying* (2.1) *and not identically zero modulo p for which $u_n \equiv 0 \pmod{p}$ for $k-1$ arbitrary values of n.*

Proof. Write $(u_n)=(c_0w_n+c_1w_{n+1}+\cdots+c_{k-1}w_{n+k-1})$, where (w_n) is the unit sequence and the c's are to be determined by the following $k-1$ congruences:

$$u_{n_1}\equiv u_{n_2}\equiv\cdots\equiv u_{n_{k-1}}\equiv 0 \pmod{p}. \tag{13.4}$$

These are $k-1$ linear homogeneous congruences in k variables, and so there must exist a solution in which not all the c's vanish, and in turn (u_n) does not vanish identically. To know the exact number of solutions, we must know the rank of the w-matrix involved, and in general this is not known.

The following notation and terminology are almost exactly those of Ward [12], p. 170:

τ = period; μ = reduced period;
e = exponent to which basic multiplier m belongs mod p;
$\phi(f)$ = number of residues prime to $p, f(x)$;
κ = number of blocks.

$$\begin{aligned} x^{\mu} &\equiv m(p, f(x)); \\ \tau &= e\mu; \qquad et = p-1; \\ \phi(f) &= (p-1)\mu\kappa = t\tau\kappa. \end{aligned} \tag{13.5}$$

THEOREM 13.4. *Let $0<a<\mu$. Then among the representative reduced cycles, one from each block, there are precisely $(p^{k-2}-1)/(p-1)$ pairs of zeros in positions differing by a.*

Proof. In an unordered cycle there are as many pairs of zeros differing in position by a as there are ordered cycles made from it with zeros u_n and u_{n+a}, n fixed. In the block there will be $p-1$ times as many pairs of zeros differing in position by a as there are in a representative reduced cycle. Hence we must find the number of solutions of

$$u_n\equiv u_{n+a}\equiv 0 \pmod{p} \tag{13.6}$$

not vanishing identically and then divide by $p-1$. If we write $(u_n)=(c_0w_n+\cdots+c_{k-1}w_{n+k-1})$, where (w_n) is the unit sequence, then the number of solutions of (13.6) will depend on the rank of the matrix

$$\begin{pmatrix} w_n, & w_{n+1}, & \cdots, & w_{n+k-1} \\ w_{n+a}, & w_{n+a+1}, & \cdots, & w_{n+a+k-1} \end{pmatrix}. \tag{13.7}$$

The rank of this matrix is two. It cannot be zero, for k consecutive terms of the unit sequence may not vanish. It cannot be one, for then the two rows would be proportional and a would be a multiple of the reduced period. Hence (13.6) has p^{k-2} solutions, of which one vanishes identically.

We are now in a position to prove the very interesting theorem:

THEOREM 13.5. *If $f(x)$ is irreducible modulo p, and b_i is the number of zeros in a reduced cycle of the ith block (the blocks having been appropriately ordered), then we have the following equations**

$$\begin{aligned} \sum b_i &= (p^{k-1}-1)/(p-1), \\ \sum (b_i^2 - b_i) &= (\mu-1)(p^{k-2}-1)/(p-1), \\ \sum b_i b_{i+r} &= (p^{k-2}-1)/(p-1), \qquad r = 1, \cdots, \kappa-1. \end{aligned} \tag{13.8}$$

LEMMA 1. *If $f(x)$ is irreducible modulo p, then*

$$e = (\tau, p-1). \tag{13.9}$$

Let $s=(\tau, p-1)$. From (13.5) both τ and $p-1$ are multiples of e, and hence s is a multiple of e. Let $\tau=ws$, where w must be a divisor of μ. From the congruence for the period we have

$$(x^w)^s \equiv 1 \pmod{p, f(x)} \tag{13.10}$$

or, writing $x^w=z$,

$$z^s \equiv 1 \pmod{p, f(x)}. \tag{13.11}$$

Now as s is a divisor of $p-1$ this congruence must have s rational solutions, namely those of $z^s \equiv 1 \pmod{p}$. But since $f(x)$ is irreducible, the residues modd $p, f(x)$ form a Galois field and no equation can have more solutions than its degree. Hence all the possible values of z in (13.11) must be rational and we have

$$x^w \equiv \text{rational} \pmod{p, f(x)}; \tag{13.12}$$

whence w must be a multiple of μ. But as w is also a divisor of μ, we must have

$$w = \mu, \qquad e = r = (\tau, p-1), \tag{13.13}$$

and the lemma is proved.

From the properties of the group of residues modulo $p, f(x)$ we may find a primitive root $y \pmod{p, f(x)}$ such that

$$y^{\kappa t} \equiv x \pmod{p, f(x)}. \tag{13.14}$$

The period of y is p^k-1 and by application of Lemma 1 to y-cycles (Theorem 13.2) the reduced y-period is $\mu\kappa=(p^k-1)/(p-1)$. There is but one y-cycle

* In this theorem it is tacitly assumed that the recurrence is at least of third order. There can be no zeros in a cycle of a first order recurrence. In a second order sequence there is only one reduced cycle which contains any zeros, and this contains only one zero. A partial criterion for the appearance of a multiple of p in a second order sequence was given in Ward [13] and a complete criterion in Hall [5].

and the x-cycles, cosets of $\{x\}$ in G are obtained by starting from any term of the y-cycle and taking every κtth term after that. The reduced x-cycles form the terms $y^{n\kappa t+i}$, where $n=0, 1, \cdots, \mu-1$, and different (ordered) reduced cycles will correspond to different values of i.

LEMMA 2. *Two reduced x-cycles corresponding to i_1 and i_2 will belong to the same block if and only if $i_1 \equiv i_2 \pmod{\kappa}$.*

The number of blocks is κ, and hence it is sufficient to show that multiplying a reduced x-cycle by y^κ takes it into another reduced cycle of the same block. From (13.9) and (13.13)

$$e = (\tau, p-1) = (\mu e, et), \tag{13.15}$$

whence

$$(\mu, t) = 1. \tag{13.16}$$

Hence the equation $\mu s + tq = 1$ has integral solutions s and q. Multiplying we get $\mu\kappa s + \kappa t q = \kappa$. Thus $y^\kappa = y^{\mu\kappa s} y^{\kappa t q}$. But $\mu\kappa$ is the reduced period of y and hence

$$y^{\mu\kappa} \equiv g \pmod{p, f(x)}, \tag{13.17}$$

where g is rational. From (13.14) and (13.17) we now have

$$y^\kappa \equiv g^s x^q \pmod{p, f(x)}, \tag{13.18}$$

and multiplication of a reduced x-cycle by such a quantity takes it into another reduced cycle of the same block.

In consequence of this lemma, we may choose as representative reduced cycles, one from each block, the terms

$$y^{n\kappa t+i}, \qquad n = 0, 1, \cdots, \mu-1; \; i = 0, 1, \cdots, \kappa-1. \tag{13.19}$$

LEMMA 3. *There are as many zeros in the terms $y^{n\kappa+i}$ as there are in the terms $y^{n\kappa t+i}$, where $n=0, 1, \cdots, \mu-1$.*

In any cycle, two terms which differ in position by a multiple of the reduced period are both zeros or neither. Applying this to the y-cycle, as $n=0, 1, \cdots, \mu-1$, $n\kappa t+i$ and $n\kappa+i$ take on the same residues modulo $\mu\kappa$ and hence must contain the same number of zeros.

From Lemma 3, b_i is the number of zeros in the terms $y^{n\kappa+i}$, $(n=0, \cdots, \mu-1)$. From Lemma 2 we shall have enumerated one cycle from each block if we take $i=0, 1, \cdots, \kappa-1$. Together all these terms form a single reduced y-cycle. By Theorem 13.4 in this reduced y-cycle there will be $(p^{k-2}-1)/(p-1)$ pairs of zeros differing by a in position for every a from 1 to $\mu\kappa-1$. Two zeros

belong to the same x-cycle if and only if they differ in position in the y-cycle by a multiple of κ. Among the differences $1, \cdots, \mu\kappa-1$ there are $\mu-1$ multiples of κ. Hence $(\mu-1)(p^{k-2}-1)/(p-1)$ differences arise from pairs of zeros in the same x-cycle. But from the cycle of the ith block there are $b_i(b_i-1)$ differences. The second of equations (13.8) arises from equating the two methods of enumerating these differences. Two zeros of the y-cycle differing in position by a number congruent to r modulo κ, will be in blocks i and $i+r$. Enumerating these differences we obtain the final equations of (13.8). The first of equations (13.8) states merely that there are a total of $(p^{k-1}-1)/(p-1)$ zeros in the reduced y-cycle or $p^{k-1}-1$ zeros in the complete y-cycle. Or in other words, there are $p^{k-1}-1$ residues modulo p, $f(x)$ whose leading coefficient vanishes, excluding the residue zero.

If we consider the third order recurrence

$$u_{n+3} = 3u_{n+2} - 2u_{n+1} - u_n \tag{13.20}$$

and its cycles modulo 11, we find that $\mu=19$ and $\kappa=7$. The b's in order are 3, 3, 1, 3, 1, 1, 0 and it is easily verified that these numbers satisfy equations (13.8). Similarly for the fourth order recurrence

$$u_{n+4} = -u_{n+3} + u_{n+2} + u_{n+1} - u_n \tag{13.21}$$

considered modulo 5, we find that $\mu=13$ and $\kappa=12$, and that the b's in order are 3, 4, 5, 2, 3, 0, 3, 2, 1, 4, 3, 1.

In both these instances there are sequences satisfying the recurrence which contain no multiples of p, namely those belonging to the block whose $b=0$. It is interesting to ask if there must be zeros in every cycle if the reduced period is large enough. This is answered by the following theorem:

THEOREM 13.6. *A sequence will contain multiples of a prime p if $f(x)$ is irreducible modulo p and $\mu > p^{k/2}$.*

This theorem is proved by obtaining $\mu < p^{k/2}$ as a consequence of the assumption that some $b=0$. Suppose that one of the b's is zero. We observe that if the sum of a fixed number of real numbers is fixed, then the sum of their squares is a minimum when all are equal. Hence

$$\sum b_i^2 \geqq (\kappa - 1)\left(\frac{\sum b_i}{\kappa - 1}\right)^2. \tag{13.22}$$

Substituting the known value of $\sum b_i$ and $\sum b_i^2$ from (13.8), we obtain after some calculation

$$\mu \leqq p^{k/2} - (p-1) + \frac{(p-1)(p^{(k-2)/2}-1)}{p^{k-2}-1}; \tag{13.23}$$

whence certainly μ is less than $p^{k/2}$.

BIBLIOGRAPHY

1. R. D. Carmichael. *On sequences of integers defined by recurrence relations*, Quarterly Journal of Mathematics, vol. 48 (1920), pp. 343–372.

2. M. Deuring. *Algebren*, Ergebnisse der Mathematik, Berlin, Springer, 1935.

3. H. T. Engstrom. *On sequences defined by linear recurrence relations*, these Transactions, vol. 33 (1931), pp. 210–218.

4. M. Hall. *Divisibility sequences of third order*, American Journal of Mathematics, vol. 58 (1936), pp. 577–584.

5. ———, *Divisors of second order sequences*, Bulletin of the American Mathematical Society, vol. 43 (1937), pp. 78–80.

6. D. H. Lehmer. *An extended theory of Lucas' functions*, Annals of Mathematics, (2), vol. 31 (1930), pp. 419–448.

7. E. Lucas. *Théorie des fonctions numériques simplement périodiques*, American Journal of Mathematics, vol. 1 (1878), pp. 184–239, 289–321.

8. O. Ore. *Contributions to the theory of finite fields*, these Transactions, vol. 36 (1934), pp. 243–274.

9. F. K. Schmidt. *Zur Zahlentheorie in Körpern von der Charakteristik p*, Erlangen Sitzungsberichte, vol. 58 (1928), pp. 159–172.

10. T. Takenouchi. *On the classes of congruent integers in an algebraic Körper*, Journal of the College of Science, Tokyo, vol. 36 (1913), pp. 1–18.

11. Morgan Ward. *The arithmetical theory of linear recurring series*, these Transactions, vol. 35 (1933), pp. 600–628.

12. ———, *The distribution of residues in a sequence satisfying a linear recursion relation*, these Transactions, vol. 33 (1931), pp. 166–190.

13. ———, *An arithmetical property of recurring series of the second order*, Bulletin of the American Mathematical Society, vol. 40 (1934), pp. 825–828.

14. ———, *Arithmetical properties of sequences in rings*, Annals of Mathematics, (2), vol. 39 (1938), pp. 210–219.

YALE UNIVERSITY,
NEW HAVEN, CONN.

RATIONAL EQUIVALENCE OF A FORM TO A SUM OF pTH POWERS*

BY

RUFUS OLDENBURGER

1. **Introduction.** In my earlier papers on p-way matrices and associated forms, I introduced new rank concepts of higher dimensional matrices with the aim of obtaining a fairly general theory of equivalence of ordinary forms and multilinear forms of arbitrary degree p, where this theory contains as a special case the well known theory of equivalence of quadratic and bilinear forms under non-singular linear transformations in a given field. In a paper in these Transactions in 1936 I gave such a development for multilinear forms. The present paper is devoted to a theory of equivalence of ordinary forms. Specifically, the problem of determining necessary and sufficient conditions for the equivalence of ordinary forms and multilinear forms of arbitrary degree to forms with "diagonal matrices" is solved in these two papers for the class of non-singular transformations in a field ϕ, where ϕ is subject to slight restrictions in the case of ordinary forms.

Throughout the present paper *we shall use the expression "a sum of pth powers" to denote a sum of the form*

$$\lambda_1 x_1^p + \cdots + \lambda_\sigma x_\sigma^p,$$

where $\lambda_1, \cdots, \lambda_\sigma$ are constant elements of a given field; the coefficients $\lambda_1, \cdots, \lambda_\sigma$ *need not be unity.*

Let $F = a_{ij\ldots m}x_i x_j \cdots x_m$ be a form of degree greater than or equal to 2 with coefficients in a field ϕ, where the matrix $A = (a_{ij\ldots m})$ of F is symmetric; that is, $a_{1\ldots12} = a_{1\ldots121} = \cdots = a_{21\ldots1}$; $a_{1\ldots122} = a_{1\ldots1212} = \cdots = a_{221\ldots1}$, for example. Let G denote a grouping of the indices of F into two classes P_1, P_2 of partitions (multipartite indices), where P_1 contains an even number, greater than or equal to 2, of partitions. Since the partitions in P_1 and P_2 play different roles in the general theory of forms, the partitions in P_1 are said to be *signant* while those of P_2 are *non-signant.* To every G there corresponds uniquely a non-negative integer D_G defined elsewhere in terms of generalized determinants associated with F. The integer D_G is not changed if ϕ is replaced by a field ψ which contains ϕ. To every grouping g of the indices

* Presented to the Society, April 10, 1936 and April 9, 1937; presented to the International Mathematical Congress, July 18, 1936; received by the editors August 19, 1937.

BIBLIOGRAPHY

1. R. D. Carmichael. *On sequences of integers defined by recurrence relations*, Quarterly Journal of Mathematics, vol. 48 (1920), pp. 343–372.

2. M. Deuring. *Algebren*, Ergebnisse der Mathematik, Berlin, Springer, 1935.

3. H. T. Engstrom. *On sequences defined by linear recurrence relations*, these Transactions, vol. 33 (1931), pp. 210–218.

4. M. Hall. *Divisibility sequences of third order*, American Journal of Mathematics, vol. 58 (1936), pp. 577–584.

5. ———, *Divisors of second order sequences*, Bulletin of the American Mathematical Society, vol. 43 (1937), pp. 78–80.

6. D. H. Lehmer. *An extended theory of Lucas' functions*, Annals of Mathematics, (2), vol. 31 (1930), pp. 419–448.

7. E. Lucas. *Théorie des fonctions numériques simplement périodiques*, American Journal of Mathematics, vol. 1 (1878), pp. 184–239, 289–321.

8. O. Ore. *Contributions to the theory of finite fields*, these Transactions, vol. 36 (1934), pp. 243–274.

9. F. K. Schmidt. *Zur Zahlentheorie in Körpern von der Charakteristik p*, Erlangen Sitzungsberichte, vol. 58 (1928), pp. 159–172.

10. T. Takenouchi. *On the classes of congruent integers in an algebraic Körper*, Journal of the College of Science, Tokyo, vol. 36 (1913), pp. 1–18.

11. Morgan Ward. *The arithmetical theory of linear recurring series*, these Transactions, vol. 35 (1933), pp. 600–628.

12. ———, *The distribution of residues in a sequence satisfying a linear recursion relation*, these Transactions, vol. 33 (1931), pp. 166–190.

13. ———, *An arithmetical property of recurring series of the second order*, Bulletin of the American Mathematical Society, vol. 40 (1934), pp. 825–828.

14. ———, *Arithmetical properties of sequences in rings*, Annals of Mathematics, (2), vol. 39 (1938), pp. 210–219.

YALE UNIVERSITY,
NEW HAVEN, CONN.

RATIONAL EQUIVALENCE OF A FORM TO A SUM OF pTH POWERS*

BY

RUFUS OLDENBURGER

1. **Introduction.** In my earlier papers on p-way matrices and associated forms, I introduced new rank concepts of higher dimensional matrices with the aim of obtaining a fairly general theory of equivalence of ordinary forms and multilinear forms of arbitrary degree p, where this theory contains as a special case the well known theory of equivalence of quadratic and bilinear forms under non-singular linear transformations in a given field. In a paper in these Transactions in 1936 I gave such a development for multilinear forms. The present paper is devoted to a theory of equivalence of ordinary forms. Specifically, the problem of determining necessary and sufficient conditions for the equivalence of ordinary forms and multilinear forms of arbitrary degree to forms with "diagonal matrices" is solved in these two papers for the class of non-singular transformations in a field ϕ, where ϕ is subject to slight restrictions in the case of ordinary forms.

Throughout the present paper *we shall use the expression "a sum of pth powers" to denote a sum of the form*

$$\lambda_1 x_1^p + \cdots + \lambda_\sigma x_\sigma^p,$$

where $\lambda_1, \cdots, \lambda_\sigma$ are constant elements of a given field; the coefficients $\lambda_1, \cdots, \lambda_\sigma$ *need not be unity*.

Let $F = a_{ij\ldots m}x_i x_j \cdots x_m$ be a form of degree greater than or equal to 2 with coefficients in a field ϕ, where the matrix $A = (a_{ij\ldots m})$ of F is symmetric; that is, $a_{1\ldots 12} = a_{1\ldots 121} = \cdots = a_{21\ldots 1}$; $a_{1\ldots 122} = a_{1\ldots 1212} = \cdots = a_{221\ldots 1}$, for example. Let G denote a grouping of the indices of F into two classes P_1, P_2 of partitions (multipartite indices), where P_1 contains an even number, greater than or equal to 2, of partitions. Since the partitions in P_1 and P_2 play different roles in the general theory of forms, the partitions in P_1 are said to be *signant* while those of P_2 are *non-signant*. To every G there corresponds uniquely a non-negative integer D_G defined elsewhere in terms of generalized determinants associated with F. The integer D_G is not changed if ϕ is replaced by a field ψ which contains ϕ. To every grouping g of the indices

* Presented to the Society, April 10, 1936 and April 9, 1937; presented to the International Mathematical Congress, July 18, 1936; received by the editors August 19, 1937.

of F into at least two partitions there corresponds a unique non-negative integer F_ϕ defined in terms of factorization properties of the matrix A associated with F. The integer F_ϕ depends on the field ϕ in which the coefficients of F are embedded. The ranks of types D_G and F_ϕ are called determinant ranks and factorization ranks, respectively. Under non-singular linear transformations $x_i = b_{i\alpha}x'_\alpha$, $x_j = b_{j\beta}x'_\beta$, $\cdots$, $x_m = b_{m\gamma}x'_\gamma$ on F these ranks are invariant.*

To give the theory of this paper completeness, if F is a linear form $a_i x_i \neq 0$, we define for F one rank D_G and one rank F_ϕ each equal to 1.

Let $\pi_n, (p; r, s, \cdots, t)$, be the number of distinct permutations of p integers chosen from $1, \cdots, n$, where r are alike, s are alike, $\cdots$, t are alike, and $r+s+\cdots+t=p$; and where there are at least two numbers in the set $r, s, \cdots, t$. Evidently,

$$\pi_n(p; r, s, \cdots, t) = \frac{p!}{r!s!\cdots t!}.$$

Let k_{np} be the class of all integers $\pi_n, (p; r, s, \cdots, t)$, for given n, p. In the case where $p \geqq 2$ a field ϕ will be said to be (n, p)-proper if its characteristic is different from all prime factors of the numbers in the class k_{np}. Every field is said to be $(n, 1)$-proper.

The symmetric matrix $A = (a_{ij\ldots m})$ of the n-ary form $F = a_{ij\ldots m}x_i x_j \cdots x_m$ of degree p with coefficients in a field ϕ is unique if and only if the field ϕ is (n, p)-proper.

An (n, p)-proper field is evidently (m, p)-proper, where $m \leqq n$.

If F is quadratic, there is only one determinant rank and one factorization rank, and they are equal. The latter rank is trivial in this case. If F is of higher degree, the ranks of F are not always equal. A class of equality and inequality relations which exist between the ranks of F have been obtained elsewhere,† but this class is not complete.

I have proved, for example, that all determinant ranks of binary forms of any degree are equal when these ranks have exactly two partitions signant.‡

If a certain determinant rank and a certain factorization rank are equal, all of the remaining ranks of F are equal. *The leading contribution of the present paper is Theorem 2 which states that the ranks of an n-ary form F of degree p*

---

* R. Oldenburger, *Composition and rank of n-way matrices and multilinear forms*, Annals of Mathematics, vol. 35 (1934), pp. 625–636, 645–646. We shall call this paper I.

R. Oldenburger, *Non-singular multilinear forms and certain p-way matrix factorizations*, these Transactions, vol. 39 (1936), pp. 422–456, especially p. 432. We shall call this paper II.

† See paper I and its supplement, pp. 622–657.

‡ R. Oldenburger, *Relations between ranks of a general matrix*, Annals of Mathematics, vol. 39 (1938), pp. 172–178.

with respect to an (n, p)-proper field ϕ, are equal if and only if F is equivalent under non-singular linear transformations in ϕ to a form

$$C = a_1x_1^p + \cdots + a_\sigma x_\sigma^p.$$

By the equality of ranks of a quadratic form this theorem includes the well known result* that any quadratic form is equivalent in a field of characteristic not 2 to a form C with $p=2$. Here and in what follows "*equivalent*" *will be used for "equivalent under non-singular linear transformations."*

That a multilinear form $M = a_{ij\ldots m}x_iy_j \cdots z_m$ has equal ranks for a field ϕ if and only if M is equivalent in ϕ to a form $x_1y_1 \cdots z_1 + \cdots + x_ny_n \cdots z_n$, was noted in an earlier paper.†

With every form F given by

$$a_{ij\ldots m}x_ix_j \cdots x_m, \qquad i, j, \cdots, m = 1, \cdots, n,$$

where the matrix $(a_{ij\ldots m})$ is symmetric, we can associate a multilinear form M given by

$$a_{ij\ldots m}x_iy_j \cdots z_m,$$

and uniquely determined by F. If F is equivalent under non-singular linear transformations in an (n, p)-proper field ϕ to a sum C of n pth powers, we shall say that F is *non-singular with respect to* ϕ. Theorem II of the present paper implies that *F is non-singular with respect to an (n, p)-proper field ϕ if and only if its associated multilinear form M is non-singular with respect to ϕ* (that is, M is non-singular in the sense of paper II). We have thus succeeded in generalizing the definitions of non-singularity of quadratic and bilinear forms to general forms in such a way that the above property, obviously valid for quadratic and bilinear forms, also holds for general forms F and M. The importance of my earlier Transactions paper is now more clearly brought out. The conditions obtained there for determining whether or not a given multilinear form is non-singular may be applied here to the form M to determine whether or not F is non-singular.

Since it is not always a simple matter to determine the factorization ranks of a given form F, other conditions are obtained for the equivalence of F to a sum of pth powers. It is proved that if p is odd and greater than or equal to 3 an n-ary form F is equivalent in an (n, p)-proper field ϕ to a sum of pth powers if and only if the generalized determinant $|x_ia_{ij\ldots m}|$, with $j, \cdots, m$ signant, is the product of linearly independent linear factors in ϕ, and under

* C. C. MacDuffee, *The Theory of Matrices*, Ergebnisse der Mathematik, vol. 2, part 5, 1933.

† R. Oldenburger, II, p. 432.

reduction in ϕ to $kx_1' \cdots x_n'$ the form F reduces covariantly to C. If p is even and greater than or equal to 4, the same statement applies except that $|x_i a_{ij\cdots m}|$ is replaced by $|x_i x_j a_{ijk\cdots m}|$, with $k, \cdots, m$ signant, and this determinant is a product of squares of linearly independent linear factors. Hočevar* sketched part of a proof of the property that a form factors into linear factors in the complex field if and only if it divides every third order minor of its Hessian, and he gave a method of finding these factors which involves, besides simply performed algebraic manipulations, only the solution of an algebraic equation E with an inequality side condition. Although the fact is not mentioned by Hočevar, this proof is valid only if the form has no repeated factors involving variables. Hence, assuming that the roots of E have been found subject to the side condition, we can determine directly the equivalence of a form to a sum of pth powers in the complex field in a finite number of steps. In this sense *the problem of determining whether or not a given form F is equivalent to a sum of pth powers is completely solved for the complex field.* If F is at most quaternary and of odd degree, the equation E is of the fourth degree or less and can be solved by well known methods. The same statement applies if F is n-ary, where $n \leqq 2$, and of even degree. Whether or not F is equivalent to a sum of pth powers can be determined directly, in these cases, in a finite number of steps.

There is a direct method, involving an induction process, of determining whether or not the ranks of a form F are equal. It is based upon the theorem that an n-ary form F of degree p, $(p \geqq 2)$, is equivalent to C in an (n, p)-proper field ϕ if and only if one determinant rank of F is n and the forms $F_1 = a_{ij\cdots m} x_j \cdots x_m, \cdots, F_n = a_{nj\cdots m} x_j \cdots x_m$ are simultaneously equivalent in ϕ to sums of $(p-1)$th powers. It is to be observed that $F \equiv x_1 F_1 + \cdots + x_n F_n$. The problem of simultaneous equivalence of $F_1, \cdots, F_n$ to sums of $(p-1)$th powers, where these forms are quadratic, is quite different from the problem for which these are cubic or of higher degree. This is due in part to the essential difference between transformations $x_i = b_{i\alpha} x_\alpha'$ which bring

$$C = \lambda_1 x_1^p + \cdots + \lambda_r x_r^p, \qquad \lambda_1, \cdots, \lambda_r \neq 0$$

into a similar form

$$C' = \lambda_1' x_1'^p + \cdots + \lambda_r' x_r'^p, \qquad \lambda_1', \cdots, \lambda_r' \neq 0,$$

for $p = 2$, and the transformations which bring C into C' for $p \geqq 3$. In fact, for

* Hočevar, *Sur les formes décomposables en facteurs linéaires*, Comptes Rendus de l'Académie des Sciences, vol. 138 (1904), pp. 745–747.

$p \geqq 3$ and an (n, p)-proper field ϕ the matrix $B = (b_{i\alpha})$ of the non-singular linear transformation $x_i = b_{i\alpha} x_\alpha'$, $(i, \alpha = 1, 2, \cdots, n, n \geqq r)$, bringing C into C' is of the type

$$B = \left\| \begin{matrix} B_{11} & 0 \\ B_{21} & B_{22} \end{matrix} \right\|, \tag{1}$$

where B_{11} is of order r, and has but one non-vanishing element in each row and column. The transformations which bring a quadratic form C into a similar form C' are thus more complicated than the transformations which bring a form C of the third degree or higher into a like form C'. There is an analogue of this for multilinear forms.*

It is to be noted that the shape of B in (1) depends on r and not on p. The solution of the general problem of equivalence in ϕ of $F_1, \cdots, F_n$ to sums of $(p-1)$th powers, $(p \geqq 4)$, depends primarily on the following statements:

1. A transformation with matrix B of (1) brings any sum of qth powers in $x_1, \cdots, x_r$ in an (n, q)-proper field into a sum of qth powers with the same number of non-vanishing coefficients.

2. For properly restricted fields the class of non-singular linear transformations bringing a form C with $p \geqq 3$ into a like form C' is identical with the class of transformations which brings a set of forms $L, M, \cdots, Q$ into a like set, where $L, M, \cdots, Q$ are sums of qth powers, $(q \geqq 3)$, in $x_1, \cdots, x_r$ and have the property that there exists for each variable x_σ in the set $x_1, \cdots, x_r$ at least one form in the set having a non-vanishing coefficient for x_σ^q.

3. If an n-ary form $F = a_{ij\ldots m} x_i x_j \cdots x_m$ is equivalent in an (n, p)-proper field ϕ to a sum of pth powers under transformation with matrix B of (1), the terms in F involving only $x_{r+1}, \cdots, x_n$ are equivalent in ϕ, under non-singular linear transformations on $x_{r+1}, \cdots, x_n$, to a sum of pth powers.

By these and other properties, the problem of equivalence of a form F to a sum of pth powers is solved by treating consecutively the equivalence of sets of subforms associated with F, where the forms in these sets are of lower degree than F. For the sake of brevity some of the more involved parts of the theory will be omitted.

In the study of non-singular multilinear forms carried through in an earlier paper, difficulties arose† in the process of determining the non-singularity of a given form. The analogue for multilinear forms of the above induction process avoids these difficulties entirely, so that it is a relatively simple

* R. Oldenburger, II, pp. 425–427, 442–443.

† R. Oldenburger, II, p. 431.

matter to determine whether or not a given form is non-singular. Since the analogue involves few new features, its presentation will be omitted.

The induction process when applied to binary forms leads to conditions, like those mentioned above, involving the generalized determinants $|x_i a_{ij\cdots m}|$, and $|x_i x_j a_{ijk\cdots m}|$, except that the conditions are much stronger in that the determinants are replaced by ordinary second order 2-way determinants, and no distinction is made between even and odd values of p. Since the binary case involves at most the determination of whether or not $(\epsilon)^{1/2}$ is in ϕ, given that ϵ is in ϕ, we may consider the problem of equivalence of a binary form of any degree to be completely solved for a $(2, p)$-proper field in the sense that this equivalence can be recognized by simple direct steps. This development can be used to give necessary and sufficient conditions for the equivalence of an algebraic equation $P(x)=0$ to the equation

$$y^p - A = 0$$

under linear fractional transformations in a $(2, p)$-proper field ϕ. This problem was considered for a finite field $GF(p)$ by Brahana.*

Bronowski† considered the problem of the equivalence under transformations, not necessarily non-singular, of a form of the pth degree to a form of the type

$$\sum_i y_i^p + \sum_i \lambda_i Y_i^p.$$

Bronowski translated the problem into one in geometry which has not been solved.

Sylvester proved‡ that a fairly broad class of binary forms of degree p can be written in the complex field as sums of pth powers of linear forms. These linear forms are, however, in general not linearly independent.

It will be proved elsewhere that *every form F with symmetric matrix can be written in a field with p or more elements as a sum*

$$a_1 L_1^p + \cdots + a_r L_r^p,$$

where r is finite, $L_1, \cdots, L_r$ are linear forms, and $a_1, \cdots, a_r$ are in the given field. For a field with less than p elements this representation is not in general possible. The forms which can be represented as above, where $L_1, \cdots, L_r$

* Brahana, *Note on irreducible quartic congruences*, these Transactions, vol. 38 (1935), pp. 395–400.

† Bronowski, *The sum of powers of a canonical expression*, Proceedings of the Cambridge Philosophical Society, vol. 29 (1933), pp. 69–82.

‡ J. J. Sylvester, Philosophical Magazine, 1851, p. 94.

are linearly independent, are the forms with which the present paper is concerned. In this case *r takes on its minimum value* for the class of forms of degree p in r variables, where the forms in this class cannot be written by means of non-singular linear transformations in terms of less than r variables, and these forms have symmetric matrices.

It is a significant thing that the representability of a form of odd degree in a field ϕ by a sum of powers

$$aL^p + bM^p + \cdots + dN^p$$

of linearly independent linear forms $L, M, \cdots, N$ depends directly on the factorability of a second form in ϕ into a product $kRS \cdots T$, of linearly independent linear factors $R, S, \cdots, T$.

The generalized *matrix method of approach* used here is a new one in the theory of forms of degree higher than quadratic.

Throughout the present paper *when we equate two forms of degree p we will assume that the field of coefficients has $p+1$ or more elements*, so that equality of forms implies equality of the corresponding coefficients.

2. **Ranks of general matrices and forms.** Hitchcock* studied certain ranks of a p-way matrix. In another paper I proved† that generalized determinant minors of the product

$$\left(\sum_m a_{ij\cdots rm} b_{m\alpha\cdots\beta} \right)$$

of two matrices $A = (a_{ij\cdots m})$ and $B = (b_{m\alpha\cdots\beta})$ are not always sums of products of determinant minors of A and B, but rather sums of products of determinant minors of matrices called "derivates" associated with A and B. In terms of determinant minors of A and derivates of A, I defined determinant ranks of A for any grouping of the indices into partitions and allowable signancy‡ of these partitions, and proved the invariance§ of these ranks under non-singular linear transformations on the form $a_{ij\cdots m}x_iy_j \cdots z_m$ associated with A. The definitions of a few determinant ranks essential to the argument will be given explicitly.

The matrix $A = (a_{ij\cdots m})$, $(i, j, \cdots, m = 1, \cdots, n)$, is said to be of *order n*. We shall assume in what follows that A is symmetric. The ranks of a form $F = a_{ij\cdots m}x_ix_j \cdots x_m$ are the ranks of its associated matrix A.

* F. L. Hitchcock, *Multiple invariants and generalized rank of a p-way matrix or tensor*, Journal of Mathematics and Physics, vol. 7 (1927), pp. 40–79.

† R. Oldenburger, I, p. 632.

‡ For a discussion of signancy of generalized determinants see L. H. Rice, *Introduction to higher determinants*, Journal of Mathematics and Physics, vol. 9 (1930), p. 48.

§ R. Oldenburger, I, pp. 633–635, 645–646.

The simplest determinant rank of the matrix A is the ordinary rank of the 2-way matrix $(a_{i\tau})$ obtained from A by letting i be the index of the rows of $(a_{i\tau})$, and τ the index of the columns of $(a_{i\tau})$, where τ is the partition of indices $j \cdots lm$ ranging over the values $1 \cdots 11$, $1 \cdots 12, \cdots, 1 \cdots 1n$, $1 \cdots 21, 1 \cdots 22, \cdots, 1 \cdots 2n, \cdots, n \cdots nn$, where n is the order of A. This rank will be called the *principal determinant rank of* A. It was used by Mayer in a paper in these Transactions.*

Let the indices of $A=(a_{ij\cdots m})$ be grouped into partitions $\rho, \sigma, \cdots, \tau$. The minimum value of ϵ for which A can be written in the form

$$(a_{ij\cdots m}) \equiv \left(\sum_{\alpha=1}^{\epsilon} b_{\alpha\rho} c_{\alpha\sigma} \cdots d_{\alpha\tau} \right),$$

where $b_{\alpha\rho}, \cdots, d_{\alpha\tau}$ are in a given field ϕ, is called the $(\rho\sigma \cdots \tau)$-factorization rank of A with respect to ϕ. The number ϵ is always finite. The $(ij \cdots m)$-factorization rank of A with respect to ϕ is called the *principal factorization rank*† of A with respect to ϕ. That this rank depends on ϕ is evident from the following example.

The form x^3-3xy^2 has a matrix $A=(a_{ijk})$ for which $a_{111}=1$, $a_{122}=a_{212}=a_{221}=-1$ and all other elements vanish. The matrix A can be written as

$$\left(\sum_{\alpha=1}^{2} b_{\alpha i} b_{\alpha j} b_{\alpha k} \right),$$

where

$$(b_{\alpha i}) = \begin{Vmatrix} \dfrac{1}{(2)^{1/3}} & \dfrac{I}{(2)^{1/3}} \\ \dfrac{1}{(2)^{1/3}} & \dfrac{-I}{(2)^{1/3}} \end{Vmatrix}, \qquad I = (-1)^{1/2}.$$

Since the principal determinant rank of A is 2, the principal factorization rank is at least 2. Therefore, the principal factorization rank of A with respect to the complex field is 2. In the complex field x^3-3xy^2 is equivalent to x^3+y^3 by the results of the present paper (Theorem I). In the field of reals the principal factorization rank of A is at least 3 since otherwise x^3-3xy^2 would be equivalent in the field of reals to $\lambda x^3+\mu y^3$ by Theorem I of this paper. It follows from Theorem IIIa of the present paper that these forms are not equivalent for this field.

---

* W. Mayer, *Die Differentialgeometrie der Untermannigfaltigkeiten des R_n konstanter Krümmung*, these Transactions, vol. 38 (1935), pp. 274–310.

† R. Oldenburger, *On arithmetic invariants of binary cubic and binary trilinear forms*, Bulletin of the American Mathematical Society, vol. 42 (1936), pp. 871–873.

If the principal factorization and determinant ranks of $F = a_{ij\ldots m}x_i x_j \cdots x_m$ are equal for a field for which $(a_{ij\ldots m})$ is unique, by the definitions of these ranks the form F is equivalent at once, for some n, to a like form for which

$$(a_{ij\ldots m}) = \left(\sum_{\alpha=1}^{n} b_{\alpha i}c_{\alpha j}\cdots d_{\alpha m}\right),$$

where $(b_{\alpha i}), (c_{\alpha j}), \cdots, (d_{\alpha m})$ are non-singular matrices of order n. The associated multilinear form $M = a_{ij\ldots m}x_i y_j \cdots z_n$ is equivalent to $M' = x_1' y_1' \cdots z_1' + \cdots + x_n' y_n' \cdots z_n'$, the matrix of which is $(\delta_{\alpha i}\delta_{\alpha j}\cdots\delta_{\alpha m})$, where $(\delta_{\alpha i}) = (\delta_{\alpha j}) = \cdots = (\delta_{\alpha m})$ is the Kronecker delta of order n. It is a fairly simple matter to show that all of the ranks of M' equal n. Since the ranks of F are ranks of M, this proves the following lemma:

LEMMA I. *If the principal ranks of an n-ary form F of degree p are equal for an (n, p)-proper field, all ranks of F are equal.*

3. **Necessary and sufficient conditions for equality of ranks.** We shall prove the following theorem:

THEOREM I. *Let F be a given n-ary form of degree p, and let ϕ be an (n, p)-proper field. The principal determinant rank of F and the principal factorization rank of F in ϕ are equal to n if and only if F is equivalent in ϕ to a sum of n pth powers.*

It is obvious that the theorem is true for linear forms. The result for the quadratic case is well known.* We shall assume in what follows that $p \geqq 3$.

Let $F = a_{ij\ldots m}x_i x_j \cdots x_m$. From the definitions of ranks the principal ranks of F are equal to n if and only if the matrix of F can be written as

$$\left(\sum_{\alpha=1}^{n} a_{\alpha i}b_{\alpha j}\cdots d_{\alpha m}\right), \tag{2}$$

where $(a_{\alpha i}), \cdots, (d_{\alpha m})$ are non-singular.

If F is equivalent in ϕ to a sum of pth powers, there exist non-vanishing constants λ_α in ϕ and a non-singular matrix $(g_{\alpha\beta})$ with elements in ϕ such that

$$F \equiv \lambda_\alpha g_{\alpha i}g_{\alpha j}\cdots g_{\alpha m}x_i x_j \cdots x_m.$$

The matrix $A = (a_{ij\ldots m})$ of F is now of the form (2) where $(a_{\alpha i}) = (\lambda_\alpha g_{\alpha i})$,

$$(b_{\alpha j}) = (g_{\alpha j}), \cdots, (d_{\alpha m}) = (g_{\alpha m}), \quad \alpha \text{ not summed.}$$

Since $(g_{\alpha\beta})$ is non-singular, $(a_{\alpha i}), \cdots, (d_{\alpha m})$ are non-singular.

* C. C. MacDuffee, *The Theory of Matrices.*

Conversely, assume that the matrix A of F is in the form (2). Let $(A^{i\beta})$ be the inverse of $(a_{\alpha i})$. Applying the transformation

$$x_i = A^{i\beta}x'_\beta$$

we obtain a form whose matrix* is of the form

$$(\delta_{\alpha\beta}b'_{\alpha j}\cdots d'_{\alpha m}),$$

where $(\delta_{\alpha\beta})$ is a Kronecker delta. It is therefore no restriction on the generality of the method to assume at the start that $(a_{\alpha i})=(\delta_{\alpha i})$. We shall prove that the form F is then a sum of pth powers.

Write $A=(\delta_{\alpha i}b_{\alpha j}c_{\alpha k}q_{\alpha l}\cdots e_{\alpha t}d_{\alpha m})$, and $B=(b_{\alpha j}), C=(c_{\alpha k}), \cdots, D=(d_{\alpha m})$. Assume that $b_{11}=0$. Since B is non-singular, $(b_{12},\cdots,b_{1n})\neq(0,\cdots,0)$. The symmetry of A and the invariance of symmetry imply that

$$\begin{aligned} b_{12}c_{1k}\cdots e_{1t}d_{11} &= b_{11}c_{1k}\cdots e_{1t}d_{12},\\ &\cdots\cdots\cdots\cdots\cdots\cdots\\ b_{1n}c_{1k}\cdots e_{1t}d_{11} &= b_{11}c_{1k}\cdots e_{1t}d_{1n}, \end{aligned} \tag{3}$$

for all values of $k,\cdots,t$. The right-hand members of (3) vanish. The direct product $H=C\times\cdots\times E$ of $C,\cdots,E$ is the matrix $(c_{jk}\cdots e_{mt})$ whose elements are the possible products of elements of $C,\cdots,E$.

Display H as an ordinary 2-way matrix $(h_{\rho\sigma})$, where ρ is the partition $j\cdots m$ and σ is the partition $k\cdots t$. By a lemma proved elsewhere† the determinant of $(h_{\rho\sigma})$ is a product of powers of determinants of $C,\cdots,E$. Since these are non-singular, H is non-singular as displayed. Since the products $c_{1k}\cdots e_{1t}$ are the elements of a row of $(h_{\rho\sigma})$, and since this matrix is non-singular, these products cannot all vanish. It follows from (3) that $d_{11}=0$. By the same argument $c_{11}=\cdots=e_{11}=0$.

The symmetry of A implies that

$$b_{11}\cdots e_{11}d_{1m}=b_{m1}\cdots e_{m1}d_{m1} \tag{4}$$

for every m. Since the left member of (4) vanishes, for every m some quantity in the set $b_{m1},\cdots,d_{m1}$ vanishes. For $m=2$ it is no restriction to take this to be b_{21}. Since

$$b_{21}c_{2j}q_{21}\cdots d_{21}=b_{2j}c_{21}\cdots d_{21}$$

---

* The matrix can then be written as $(b_{\beta j}\cdots d_{\beta m})$ where β is not summed. Necessary and sufficient conditions for the factorability of a matrix in this form were given in II, p. 452.

† R. Oldenburger, I, p. 625; II, p. 442.

for every j, and since by the non-singularity of B $(b_{22}, \cdots, b_{2n}) \neq (0, \cdots, 0)$, it follows that

$$c_{21} \cdots d_{21} = 0.$$

Since one of the quantities $c_{21}, \cdots, d_{21}$ vanishes, it is no restriction to take this to be c_{21}. Since

$$b_{21}c_{2j}q_{2i}f_{21} \cdots d_{21} = b_{2i}c_{2j}q_{21}f_{21} \cdots d_{21}, \tag{5}$$

for all i, j, and $b_{2i}c_{2j}$ cannot be zero for all i, j (by the non-singularity of B, C), and since the left-hand member of (5) vanishes,

$$q_{21} \cdots d_{21} = 0.$$

It is seen by induction that $q_{21} = \cdots = d_{21} = 0$. By the same argument we prove that

$$b_{m1} = \cdots = d_{m1} = 0$$

for every m. Since $B, \cdots, D$ are non-singular, the elements $b_{11}, \cdots, d_{11}$ cannot be zero. It follows that no diagonal elements of $B, \cdots, D$ vanish.

We shall now prove that the non-diagonal elements of $B, \cdots, D$ vanish.

If some non-diagonal element of B vanishes, we may take it to be b_{12}. By the symmetry of A

$$b_{11}c_{12}q_{11} \cdots d_{11} = b_{12}c_{11}q_{11} \cdots d_{11} = b_{21}c_{21}q_{21} \cdots d_{21}; \tag{6}$$

whence $c_{12} = 0$. Hence $q_{12} = \cdots = d_{12} = 0$. By (6) some element in the set $b_{21}, \cdots, d_{21}$ vanishes. It is no restriction on the generality of the method to take this to be b_{21}. From

$$b_{21}c_{22}q_{22} \cdots d_{22} = b_{22}c_{21}q_{22} \cdots d_{22}$$

it follows that $c_{21} = 0$. Hence $q_{21} = \cdots = d_{21} = 0$. We have proved that if $b_{rs} = 0$ for a given r, s, where $r \neq s$, then $b_{sr}, c_{rs}, c_{sr}, \cdots, d_{rs}, d_{sr} = 0$.

If elements other than b_{12} in the first row of B are zero, it is no restriction to take them to be $b_{13}, \cdots, b_{1r}$ and to assume that $b_{1,r+1}, \cdots, b_{1n} \neq 0$. Then

$$b_{1j} = c_{1j} = \cdots = d_{1j} = b_{j1} = c_{j1} = \cdots = d_{j1} = 0, \qquad j = 2, \cdots, r,$$

and

$$b_{1j}, b_{j1}, c_{1j}, c_{j1}, \cdots, d_{1j}, d_{j1} \neq 0, \qquad j = r+1, \cdots, n.$$

By the symmetry of A,

$$b_{r+1,1}c_{r+1,1} \cdots e_{r+1,1}d_{r+1,j} = b_{1j}c_{11} \cdots e_{11}d_{1,r+1} = 0,$$

for $j = 2, \cdots, r$. Hence

$$d_{r+1,2} = \cdots = d_{r+1,r} = 0.$$

It follows at once that $d_{\alpha m}=0$, for $\alpha=r+1, \cdots, n$; $m=2, \cdots, r$, and $\alpha=2, \cdots, r$; $m=r+1, \cdots, n$. Evidently, also $b_{\alpha m}= \cdots =e_{\alpha m}=0$ for the same pairs of ranges of α and m.

We have proved that

$$(7) \qquad B = \begin{Vmatrix} b_{11} & 0 \cdots 0 & b_{1,r+1} & \cdots b_{1n} \\ 0 & b_{22} \cdots b_{2r} & 0 & \cdots 0 \\ \cdot & \cdot \quad \cdots \cdot & \cdot & \cdots \cdot \\ 0 & b_{r2} \cdots b_{rr} & 0 & \cdots 0 \\ b_{r+1,1} & 0 \cdots 0 & b_{r+1,r+1} & \cdots b_{r+1,n} \\ \cdot & \cdot \quad \cdots \cdot & \cdot & \cdots \cdot \\ b_{n1} & 0 \cdots 0 & b_{n,r+1} & \cdots b_{nn} \end{Vmatrix};$$

and $C, \cdots, D$ are of the same form. If no element of B vanishes, B is of the form (7) where $r+1=2$, and the minor

$$\begin{Vmatrix} b_{22} \cdots b_{2r} \\ \cdot \quad \cdots \cdot \\ b_{r2} \cdots b_{rr} \end{Vmatrix}$$

does not occur. The representation (7) hence includes all cases.

If some elements of the first row of

$$B^* = \begin{Vmatrix} b_{r+1,r+1} \cdots b_{r+1,n} \\ \cdot \quad \cdots \quad \cdot \\ b_{n,r+1} \quad \cdots b_{nn} \end{Vmatrix}$$

vanish, let these be $b_{r+1,r+\sigma}, \cdots, b_{r+1,n}$, $(\sigma \geqq 2)$. By the above reasoning $b_{r+\sigma,1}$ vanishes. Since this contradicts an earlier assumption, it follows that no elements of B^* are zero.

By simultaneous interchanges of the rows and columns of $B, \cdots, D$, corresponding to permutation of the variables of F, we can bring B into the form

$$B = \begin{Vmatrix} b_{11} \cdots b_{1s} & 0 & \cdots 0 \\ \cdot \quad \cdots \cdot & \cdot & \cdots \cdot \\ b_{s1} \cdots b_{ss} & 0 & \cdots 0 \\ 0 \quad \cdots 0 & b_{s+1,s+1} & \cdots b_{s+1,n} \\ \cdot \quad \cdots \cdot & \cdot & \cdots \cdot \\ 0 \quad \cdots 0 & b_{n,s+1} & \cdots b_{nn} \end{Vmatrix}, \qquad s = n - r + 1,$$

and $C, \cdots, D$ into similar forms where the elements of

$$\left\| \begin{matrix} b_{11} \cdots b_{1s} \\ \cdot \quad \cdots \quad \cdot \\ b_{s1} \cdots b_{ss} \end{matrix} \right\|, \cdots, \left\| \begin{matrix} d_{11} \cdots d_{1s} \\ \cdot \quad \cdots \quad \cdot \\ d_{s1} \cdots d_{ss} \end{matrix} \right\|$$

are all non-zero.

Applying the above argument to the set of minors

$$\left\| \begin{matrix} b_{s+1,s+1} \cdots b_{s+1,n} \\ \cdot \quad \cdots \quad \cdot \\ b_{n,s+1} \quad \cdots b_{nn} \end{matrix} \right\|, \cdots, \left\| \begin{matrix} d_{s+1,s+1} \cdots d_{s+1,n} \\ \cdot \quad \cdots \quad \cdot \\ d_{n,s+1} \quad \cdots d_{nn} \end{matrix} \right\|$$

we conclude that $B, \cdots, D$ can be written in the form

$$(8) \qquad B = \left\| \begin{matrix} B_1 & 0 & \cdots 0 \\ 0 & B_2 \cdots 0 \\ \cdot & \cdot & \cdots \cdot \\ 0 & 0 & \cdots B_w \end{matrix} \right\|, \cdots, \quad D = \left\| \begin{matrix} D_1 & 0 & \cdots 0 \\ 0 & D_2 \cdots 0 \\ \cdot & \cdot & \cdots \cdot \\ 0 & 0 & \cdots B_w \end{matrix} \right\|,$$

where the elements of the minors $B_1, \cdots, B_w, \cdots, D_1, \cdots, D_w$ are not zero, and $B_i, \cdots, D_i$ are minors of the same order for every i.

Assuming that $B_1, \cdots, B_w$ are not all of order 1, we may write $B, \cdots, D$ in the form (8), where $B_1, \cdots, D_1$ are of order $R \geqq 2$.

The determinant of B is

$$|B| = |B_1| \cdot |B_2| \cdot \cdots \cdot |B_w|.$$

We can write

$$|B_1| = \frac{1}{c_{12}c_{21}q_{11} \cdots d_{11}q_{21} \cdots d_{21}} \begin{vmatrix} b_{11}c_{12}q_{11} \cdots d_{11} & b_{12}c_{12}q_{11} \cdots d_{11} \cdots b_{1R}c_{12}q_{11} \cdots d_{11} \\ b_{21}c_{21}q_{21} \cdots d_{21} & b_{22}c_{21}q_{21} \cdots d_{21} \cdots b_{2R}c_{21}q_{21} \cdots d_{21} \\ b_{31} & b_{32} \quad \cdots \quad b_{3R} \\ \cdot & \cdot \quad \cdots \quad \cdot \\ b_{R1} & b_{R2} \quad \cdots \quad b_{RR} \end{vmatrix},$$

since the denominator is not zero.

By the symmetry of A

$$\begin{aligned} b_{11}c_{12}q_{11} \cdots d_{11} &= b_{21}c_{21}q_{21} \cdots d_{21}, \\ b_{12}c_{12}q_{11} \cdots d_{11} &= b_{22}c_{21}q_{21} \cdots d_{21}, \\ &\cdots\cdots\cdots\cdots\cdots, \\ b_{1R}c_{12}q_{11} \cdots d_{11} &= b_{2R}c_{21}q_{21} \cdots d_{21}, \end{aligned}$$

whence the first two rows of the above determinant are identical. Hence

$|B_1|=0$, whence B is singular. This gives a contradiction. Minors of (8) are of order 1 and

$$B=\begin{Vmatrix} b_{11} & 0 & \cdots & 0 & 0 \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & \cdots & 0 & b_{nn} \end{Vmatrix}, \cdots, \quad D=\begin{Vmatrix} d_{11} & 0 & \cdots & 0 & 0 \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & \cdots & 0 & d_{nn} \end{Vmatrix}.$$

Now

$$F=\delta_{ai}b_{aj}\cdots d_{am}x_ix_j\cdots x_m=b_{11}\cdots d_{11}x_1^p+\cdots+b_{nn}\cdots d_{nn}x_n^p,$$

which completes the proof of the theorem.

The above argument is valid if there are only two matrices in the set $B, \cdots, D$.

By Lemma I we can state Theorem I in the following form:

THEOREM Ia. *Let F and ϕ be given as in Theorem* I. *The ranks of a form F of degree p are equal in ϕ if and only if F is equivalent in ϕ to a sum of pth powers.*

If the characteristic of ϕ is 2 and F is quadratic of rank n, F is always a sum of $r\leqq n$ squares since $a_{ij}+a_{ji}=2a_{ij}=0$. This generalizes for p-ic forms, $(p\geqq 3)$. We therefore have the theorem:

THEOREM II. *Let ϕ be a field whose characteristic is a factor of all numbers in the set k_{np} of* §1, *or an (n, p)-proper field. The ranks (or the principal ranks) of an n-ary form $F=a_{ij\cdots m}x_ix_j\cdots x_m$ of degree p with respect to a field ϕ are equal if and only if F is equivalent in ϕ to a sum of pth powers.*

4. A condition for the equivalence of forms of odd degree to sums of pth powers. Let $A=(a_{ij\cdots m})$ denote a p-way matrix, where $i, j, \cdots, m=1, \cdots, n$, and p is even. Let $m_1, \cdots, m_n$ denote n distinct values of n; $j_1, \cdots, j_n$ distinct values of j; and so on. Let $(e^{i_1\cdots i_n})$ be the *generalized Kronecker delta* for which $e^{i_1\cdots i_n}=1$ if $(j_1, \cdots, j_n)$ is an even permutation of $1, \cdots, n$, and equal to -1 if $(j_1, \cdots, j_n)$ is an odd permutation of $1, \cdots, n$. The determinant of A with all indices signant is defined* to be

$$|A|=\sum e^{i_1\cdots i_n}\cdots e^{m_1\cdots m_n}a_{1j_1\cdots m_1}\cdots a_{ni_n\cdots m_n}, \tag{9}$$

where the summation is over all distinct permutations of the numbers in the sets $(j_1, \cdots, j_n), \cdots, (m_1, \cdots, m_n)$. It is of order n.

Let an n-ary form $F=a_{ij\cdots m}x_ix_j\cdots x_m$ be of odd degree p. The *i-characteristic determinant* of F is defined to be the determinant $|x_ia_{ij\cdots m}|$ with all indices $[j, \cdots, m]$ signant. If F is of degree 3, this is the Hessian of F except for a constant factor. We shall prove the following theorem:

* R. Oldenburger, I, p. 631.

THEOREM III. *An n-ary form F of odd degree $p \geqq 3$ is equivalent in an (n, p)-proper field ϕ to a sum of n pth powers, if and only if the i-characteristic determinant D of F factors in ϕ into linearly independent linear factors, and under reduction in ϕ of D to canonical form $Kx_1, \cdots, x_n$ F transforms covariantly to a sum of pth powers.*

The i-characteristic determinant of $G = \lambda_i x_i^p$ is the determinant

$$D = \left| x_i \lambda_i \delta_{ij \cdots m} \right|,$$

where $\delta = (\delta_{ij \cdots m})$ is the generalized Kronecker delta whose only non-vanishing elements are

$$\delta_{1 \cdots 1} = \cdots = \delta_{n \cdots n} = 1.$$

Evidently

$$D = Kx_1 \cdots x_n, \tag{10}$$

where $K = \lambda_1 \cdots \lambda_n$. Under non-singular transformations

$$x_i = p_{ij} x_j' \tag{11}$$

on G to give a form F, D transforms by a general theorem on products of determinants* into

$$D' = \left| P \right|^{p-1} K \prod^{i} p_{ij} x_j',$$

where $P = (p_{ij})$. The determinant D is thus a covariant of G. It follows that the i-characteristic determinant of any form F, equivalent to G, factors in ϕ into linearly independent linear factors. This is therefore a necessary condition for the equivalence in ϕ of a form to a sum of pth powers.

Assume that the i-characteristic determinant of F factors in ϕ into linearly independent linear factors $L, M, \cdots, N$. Applying the transformations

$$x_1' = L, \cdots, x_n' = N$$

to F we obtain a form F' whose i-characteristic determinant is of the form

$$Kx_1' \cdots x_n'.$$

We therefore consider only forms F where D satisfies (10). We shall prove that such a form is equivalent to a sum of pth powers if and only if it is already a sum of pth powers. If F is equivalent to a form G, then G is equivalent to F under a transformation (11) which brings (10) covariantly into $qx_1' \cdots x_n'$ for some q. For such a transformation

* R. Oldenburger, I, p. 632.

$$|P|^{p-1}K\prod^{i}(p_{ij}x_j') \equiv qx_1' \cdots x_n'.$$

By the unique factorization property, similarly used elsewhere,* the linear expressions $p_{1j}x_j', \cdots, p_{nj}x_j'$ are equal in some order to the products of $x_1', \cdots, x_n'$ by constants in ϕ. The transformations on G which bring D into $qx_1' \cdots x_n'$ are of the form

$$x_i = \alpha_j x_j', \qquad i, j = 1, \cdots, n; j \text{ not summed}, \tag{12}$$

where there is one j for every i, and conversely. Since (12) transforms G into a sum of pth powers, F is a sum of pth powers.

In the binary cubic case we have simply the following theorem:

THEOREM IIIa. *A binary cubic form $F = a_{ijk}x_ix_jx_k$ is equivalent in a field ϕ with characteristic not 3 to a sum of two cubes if and only if the i-characteristic determinant of F factors in ϕ into distinct linear factors.*

To prove this theorem it is only necessary to show that if the i-characteristic determinant D of F factors in ϕ into kx_1x_2 for some $k \neq 0$, then F is a sum of cubes. It is to be noted that ϕ is (2, 3)-proper if and only if its characteristic is not 3. Let $a_{111}=\alpha$, $a_{222}=\beta$, $a_{112}=\gamma$, $a_{122}=\delta$. Then a simple direct calculation shows that the i-characteristic determinant of F is $(\alpha\delta-\gamma^2)x_1^2+(\alpha\beta-\gamma\delta)x_1x_2 + (\gamma\beta-\delta^2)x_2^2$. If this is to be of the form kx_1x_2, we must have $\alpha\delta-\gamma^2=\gamma\beta-\delta^2 = 0$, $\alpha\beta-\gamma\delta\neq 0$. It follows readily that $\gamma=\delta=0$, $\alpha, \beta\neq 0$; thus F is of the required form.

That the factorization of the i-characteristic determinant into linearly independent linear factors is not in general enough to insure the equivalence of an n-ary form F of degree p to a sum of n pth powers follows from an example:

Example. Let $F=6x_1x_2x_3$ with matrix $A=(a_{ijk})$. The only non-vanishing elements of A are $a_{123}=a_{132}=a_{213}=a_{231}=a_{312}=a_{321}=1$. Let ϕ be a (3, 3)-proper field. The i-characteristic determinant of F is $D=2x_1x_2x_3$. Except for an integral factor it is the Hessian of F. Now D is a product of linearly independent linear factors in ϕ. Since D is in canonical form, if F is equivalent to a sum of three cubes, by Theorem III, F is such a sum. The form F is not such a sum.

The determinant $|A|$ of a p-way matrix $A=(a_{ij\ldots m})$ of order n, (p odd), where $j, \cdots, m$ are signant, is given by (9). We may have $|A| \neq 0$, even if $a_{\alpha j\ldots m} \equiv a_{\beta j\ldots m}$ for $\alpha\neq\beta$. The determinant is said to be of order n. The $(j \cdots m, i)$ derivate of A is the matrix obtained by adjoining $(n-1)$ matrices all equal to A in the direction associated with the index i in the p-space

* R. Oldenburger, *Equivalence of multilinear forms singular on one index*, Duke Mathematical Journal, vol. 2 (1936), p. 674. See also B. L. van der Waerden, *Moderne Algebra*, Springer, 1930, vol. 1, pp. 63–66.

representation of A. The *rank* $r[j \cdots m, i]$ of A and $F = a_{ij\ldots m}x_ix_j \cdots x_m$ is the upper bound of the orders of the non-vanishing determinant minors of this derivate, expanded with $j, \cdots, m$ signant.*

For an (n, n)- and (n, p)-proper field the i-characteristic determinant of F can be written as

$$\sum_{\Gamma} k_\Gamma x_{i_1} \cdots x_{i_n} \mid a_{ij\ldots m} \mid, \tag{13}$$

where the summation is over all distinct choices Γ of the values of $i_1 \cdots i_n$ from the set $1, 2, \cdots, n$; $k_\Gamma = 1/\alpha!\beta! \cdots \gamma!$, where, for example, α of the indices in Γ are equal, β other indices are equal, and $\mid a_{ij\ldots m} \mid$ is a determinant minor of the $(j \cdots m, i)$ derivate of A with $j, \cdots, m$ signant and all minors of this derivate occur in (13). There follows the theorem:

THEOREM IV. *Let ϕ be an (n, n)- and (n, p)-proper field. The i-characteristic determinant of the n-ary form $F = a_{ij\ldots m}x_ix_j \cdots x_m$ of odd degree p, $(p \geqq 3)$, is not identically zero if and only if the rank $r[j \cdots m, i]$ of F is n.*

By a theorem proved elsewhere,† the principal determinant rank of F is greater than or equal to $r[j \cdots m, i]$. Hence we have the following theorem:

THEOREM V. *Let ϕ be a field satisfying the assumptions of Theorem* IV. *If the i-characteristic determinant of an n-ary form F of odd degree greater than or equal to 3 factors into linearly independent linear factors in ϕ, the principal determinant rank of F is n.*

5. **The equivalence of forms of even degree to sums of pth powers.** With an n-ary form

$$F = a_{ij\ldots m}x_ix_j \cdots x_m$$

of even degree p, $(p \geqq 4)$, associate the determinant

$$E = \mid x_ix_ja_{ijk\ldots m} \mid$$

with all indices signant. We shall call this the *ij-characteristic determinant of F*.

We shall prove the following theorem:

THEOREM VI. *An n-ary form F of even degree p, $(p \geqq 4)$, is equivalent in an (n, p)-proper field ϕ to a sum of pth powers if and only if the ij-characteristic determinant E of F factors in ϕ into squares of linearly independent linear factors, and under reduction of E in ϕ to canonical form $Kx_1^2 \cdots x_n^2$, F transforms covariantly to a sum of pth powers.*

* R. Oldenburger, I, p. 633.

† R. Oldenburger, I, p. 641.

The proof of this theorem is very similar to that of Theorem III. The transformation from F to

$$F' = a_{ij\dots m}b_{i\alpha}b_{j\beta}\cdots b_{m\gamma}x_\alpha' x_\beta' \cdots x_\gamma'$$

under the non-singular transformation $x_i = b_{i\alpha}x_\alpha'$ corresponds, by a theorem on determinants proved elsewhere,* to the transition from E to

$$\left| a_{ijk\dots m}b_{i\alpha}b_{j\beta}x_\alpha' x_\beta' \right| \left| b_{k\gamma} \right|^{p-2}.$$

The determinant E of $G = \lambda_i x_i^p$, $(i = 1, \cdots, n; \lambda_i \neq 0)$, is $\lambda_1 \cdots \lambda_n x_1^2 \cdots x_n^2$. Evidently, the ij-characteristic determinant of a form F equivalent to G is of the form $KL^2 \cdots N^2$, where $L, \cdots, N$ are linearly independent linear factors.

Assume that the determinant E of F factors, as required in the theorem, and that F is equivalent to G. Letting $x_1' = L, \cdots, x_n' = N$ we can transform the form F into a new form for which $E = qx_1'^2 \cdots x_n'^2$. Since G is equivalent to F, there is a non-singular transformation $x_i = b_{i\alpha}x_\alpha'$ such that

$$\left| b_{i\alpha} \right|^{p-2}\lambda_1 \cdots \lambda_n(b_{1\alpha}x_\alpha')^2 \cdots (b_{n\alpha}x_\alpha')^2 \equiv qx_1'^2 \cdots x_n'^2.$$

It follows from this identity that the transformation $x_i = b_{i\alpha}x_\alpha'$ is of the form (12). Since the transformation (12) brings G into a sum of pth powers, F is a sum of pth powers.

Let $A = (a_{ijk\dots m})$ be a p-way matrix, $(p \geqq 4$ and even), of order n. The determinant of A with $k, \cdots, m$ signant is the expansion (9) except that $e^{i_1 \cdots i_n}$ does not occur. The indices i and j need not range over distinct values. The $(k \cdots m, ij)$ derivate of A is a matrix obtained from A by adjoining matrices equal to A in the directions of p-space associated with the indices i and j. The rank† $r[k \cdots m, ij]$ of A and $F = a_{ij\dots m}x_i x_j \cdots x_m$ is the upper bound of the orders of the non-vanishing determinant minors of the $(k \cdots m, ij)$ derivate of A with $k, \cdots, m$ signant in these determinants. For an (n, p)-proper field ϕ with a characteristic not equal to $2, 3, \cdots, \xi_n$, where ξ_n is a positive prime depending on n, the coefficients in the expansion of the ij-characteristic determinant of F can be written as sums of nth order determinant minors of the $(k \cdots m, ij)$ derivate of A with $k, \cdots, m$ signant. We therefore have the following theorem:

THEOREM VII. *For a field ϕ with characteristic different from $2, 3, \cdots, \xi_n$, where ξ_n is finite and large enough, the ij-characteristic determinant of an n-ary form $F = a_{ij\dots m}x_i x_j \cdots x_m$ of even degree p, $(p \geqq 4)$, is identically zero if the rank $r[k \cdots m, ij] < n$.*

* R. Oldenburger, I, p. 632.

† R. Oldenburger, I, p. 633.

Since $r[k \cdots m, ij]$ is not greater than the principal determinant rank of F,* we have the following analogue of Theorem V.

THEOREM VIII. *Let ϕ be a field with the properties of Theorem* VI. *If the ij-characteristic determinant of an n-ary form F of even degree greater than or equal to 3 factors into squares of linearly independent linear factors in ϕ, the principal determinant rank of F is n.*

6. **Method for obtaining linear factors.** Let an n-ary form $F = a_{ij\cdots m}x_i x_j \cdots x_m$, $(i, j, \cdots, m = 1, \cdots, n)$, of degree p, $(p \geqq 2)$, with elements in the complex field, be denoted by $F(x_1, \cdots, x_n)$.

Making the non-singular transformations $x_1 = x_1'$, $x_2 = \lambda x_1' + x_2'$, $x_3 = \mu x_1' + x_3'$, $\cdots$, $x_n = \nu x_1' + x_n'$ on F we obtain $F(x_1', \lambda x_1' + x_2', \cdots, \nu x_1' + x_n')$. Assuming that $F \not\equiv 0$, by the continuity of F, we have $F(1, \lambda, \cdots, \nu) \not\equiv 0$ for $\lambda, \cdots, \nu \neq 0$. Since this is the coefficient of $x_1'^p$ we have the following lemma:

LEMMA II. *A non-identically vanishing form F of degree p in $x_1, \cdots, x_n$ is equivalent in the complex field to a form for which the coefficient of x_1^p is not zero.*

Assume that F has a term in x_1^p and does not contain a repeated irreducible factor.† The last assumption implies that the resultant of F and $\partial F/\partial x_1$ involving $x_2, \cdots, x_n$ is not identically zero in $x_2, \cdots, x_n$. We can therefore choose values $a_2, \cdots, a_n$ such that

$$F(x_1, a_2, \cdots, a_n) = 0, \qquad \frac{\partial F(x_1, a_2, \cdots, a_n)}{\partial x_1} = 0$$

have no root x_1 in common.

By a theorem from implicit function theory it follows that x_1 is represented analytically by distinct power series $P_1(x_2, \cdots, x_n), \cdots, P_p(x_2, \cdots, x_n)$ at the distinct points $(\alpha_1, a_2, \cdots, a_n), \cdots, (\alpha_p, a_2, \cdots, a_n)$ for which

$$F(\alpha_i, a_2, \cdots, a_n) = 0, \quad \left[\frac{\partial F}{dx_1}\right]_{(\alpha_i, a_2, \cdots, a_n)} \not\equiv 0.$$

According to Hočevar‡ the derivatives of x_1 with respect to $x_2, \cdots, x_n$ of order higher than one involve the third order determinant minors of the Hessian H_F of F in such a way that when these minors vanish simultaneously for a point $(x_1, \cdots, x_n)$ where $\partial F/\partial x_1 \neq 0$, these derivatives also vanish. It

* R. Oldenburger, I, p. 641.

† For the case of two variables see G. A. Bliss, *Algebraic Functions*, American Mathematical Society Colloquium Publications, vol. 16, New York, 1933, pp. 29–30.

‡ Hočevar, op. cit., pp. 745–747.

follows that if F is a factor of these minors of H_F, the derivatives of x_1 of higher order than the first vanish at all points for which $F=0$, $\partial F/\partial x_1 \neq 0$. In this case the series $P_1, \cdots, P_p$ involve only terms of the first degree in $x_2, \cdots, x_n$. Except for constant multipliers, the polynomial factors of F are the expressions $x_1-P_1, \cdots, x_1-P_p$ which are linear.

Conversely, if F factors into linear factors, it can be shown at once that the third order determinant minors contain F as a factor.

Except for constant multipliers, the expressions x_i-P_i, $(i=1, \cdots, p)$, are the same as

$$x_1\left[\frac{\partial F}{\partial x_1}\right]_{p_i}+\cdots+x_n\left[\frac{\partial F}{\partial x_n}\right]_{p_i},$$

where p_i is the point $(\alpha_i, a_2, \cdots, a_n)$. These expressions are linearly independent if and only if the matrix

$$M=\left(\left[\frac{\partial F}{\partial x_i}\right]_{p_j}\right)$$

is of rank greater than or equal to p. For the D of Theorem III, and $p=n$, this condition can be stated in terms of the non-singularity of M. We have the theorem:

THEOREM IX. *Let F be a form in $x_1, \cdots, x_n$ of degree n. The form F factors in the complex field into linearly independent linear factors if and only if the following set of conditions is satisfied*:

(1) *F has no repeated factors.*

(2) *F divides the third order determinant minors of its Hessian.*

(3) *At points p_1 $(\alpha_1, a_2, \cdots, a_n), \cdots, p_n$ $(\alpha_n, a_2, \cdots, a_n)$ on $F=0$ for which there are distinct power series expansions for x_1 in terms of $x_2, \cdots, x_n$, the matrix following is non-singular*:

$$\left\|\begin{matrix}\left[\frac{\partial F}{\partial x_1}\right]_{p_1} & \cdots & \left[\frac{\partial F}{\partial x_n}\right]_{p_1}\\ \cdots & \cdots & \cdots\\ \left[\frac{\partial F}{\partial x_1}\right]_{p_n} & \cdots & \left[\frac{\partial F}{\partial x_n}\right]_{p_n}\end{matrix}\right\|.$$

7. **An induction process.** We shall prove the following theorem:

THEOREM X. *Let*

$$F=a_{ij\cdots m}x_ix_j\cdots x_m, \qquad i, j, \cdots, m=1, \cdots, n,$$

be a form of degree $p \geqq 2$, and let ϕ be (n, p)-proper. The form F is equivalent in ϕ to a sum of n pth powers if and only if

(1) *the principal determinant rank of F is n;*

(2) *the subforms $F_1 = a_{1j\ldots m}x_j \cdots x_m, \cdots, F_n = a_{nj\ldots m}x_j \cdots x_m$ are simultaneously equivalent in ϕ to a set of forms of the type*

$$G_1 = \mu_{11}y_1^{p-1} + \cdots + \mu_{1n}y_n^{p-1}, \cdots, G_n = \mu_{n1}y_1^{p-1} + \cdots + \mu_{nn}y_n^{p-1}.$$

If $p=2$ the theorem is evidently true. Assume that $p \geqq 3$. If F is equivalent to

$$G = \lambda_1 y_1^p + \cdots + \lambda_n y_n^p,$$

there exists a non-singular linear transformation

$$x_i = b_{i\alpha}y_\alpha, \qquad i, \alpha = 1, \cdots, n, \tag{14}$$

such that

$$a_{ij\ldots k}b_{i\alpha}b_{j\beta} \cdots b_{k\gamma}y_\alpha y_\beta \cdots y_\gamma = G, \tag{15}$$

where $(b_{i\alpha}) = (b_{j\beta}) = \cdots = (b_{k\gamma})$. Equation (15) implies the following matrix identity:

$$(a_{ij\ldots k}b_{i\alpha}b_{j\beta} \cdots b_{k\gamma}) = (c_{\alpha\beta\ldots\gamma}), \tag{15'}$$

where $C = (c_{\alpha\beta\ldots\gamma})$ is the matrix of G taken with all elements zero except $c_{\ldots i} = \lambda_i$, $(i = 1, \cdots, n)$. Let the inverse $(b_{i\alpha})^{-1}$ of $(b_{i\alpha})$ be denoted by $(B^{\alpha i})$. The identity now implies

$$(a_{ij\ldots k}b_{j\beta} \cdots b_{k\gamma}) = (c_{\alpha\beta\ldots\gamma}B^{\alpha i}). \tag{16}$$

Setting $i = 1, \cdots, n$ in (16) we obtain the following relations between minors of the matrices in (16):

$$\begin{aligned}(a_{1j\ldots k}b_{j\beta} \cdots b_{k\gamma}) &= (c_{\alpha\beta\ldots\gamma}B^{\alpha 1}),\\ &\cdots\cdots\cdots\cdots\cdots,\\ (a_{nj\ldots k}b_{j\beta} \cdots b_{k\gamma}) &= (c_{\alpha\beta\ldots\gamma}B^{\alpha n}).\end{aligned}$$

These equations yield

$$\begin{aligned}a_{1j\ldots k}b_{j\beta} \cdots b_{k\gamma}y_\beta \cdots y_\gamma &= c_{\alpha\beta\ldots\gamma}B^{\alpha 1}y_\beta \cdots y_\gamma,\\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots,\\ a_{nj\ldots k}b_{j\beta} \cdots b_{k\gamma}y_\beta \cdots y_\gamma &= c_{\alpha\beta\ldots\gamma}B^{\alpha n}y_\beta \cdots y_\gamma.\end{aligned} \tag{17}$$

If we let $\mu_{rs} = c_{s\ldots s}B^{sr}$, the right members of (17) become $G_1, \cdots, G_n$.

Since the principal determinant rank of G is obviously n, this rank of F is n if F is equivalent to G.

We have proved the necessity of the conditions. To prove the sufficiency we assume that the principal determinant rank of F is n, and that there exist non-singular linear transformations (14) in ϕ and a choice of μ_{ij}, $(i, j=1, \cdots, n)$, such that

$$
\begin{aligned}
& a_{1j\ldots k}b_{j\beta}\cdots b_{k\gamma}y_\beta\cdots y_\gamma = G_1, \\
& \cdots\cdots\cdots\cdots\cdots, \\
& a_{nj\ldots k}b_{j\beta}\cdots b_{k\gamma}y_\beta\cdots y_\gamma = G_n.
\end{aligned} \tag{18}
$$

By (18)

$$
a_{ij\ldots k}b_{j\beta}\cdots b_{k\gamma}x_iy_\beta\cdots y_\gamma = x_\xi G_\xi, \qquad i, \cdots, \xi = 1, \cdots, n. \tag{19}
$$

Substituting (14) in (19) we obtain

$$
a_{ij\ldots k}b_{i\alpha}b_{j\beta}\cdots b_{k\gamma}y_\alpha y_\beta\cdots y_\gamma = G_\xi b_{\xi\alpha}y_\alpha.
$$

Since A is symmetric, the matrix of $G_\xi b_{\xi\alpha}y_\alpha$ is also symmetric. Denote this matrix by $D=(d_{\alpha\beta\ldots\gamma})$. Then

$$
d_{\alpha\beta\ldots\beta} = \mu_{\xi\beta}b_{\xi\alpha}
$$

(the repeated index on the left does not indicate summation) while all other elements of D vanish. By symmetry $d_{\alpha\beta\ldots\beta}=d_{\beta\ldots\beta\alpha}$. If $\alpha\neq\beta$, then $d_{\beta\ldots\beta\alpha}=0$, whence $d_{\alpha\beta\ldots\beta}=0$ also. Hence at most $d_{1\ldots1}, \cdots, d_{n\ldots n}$ are non-zero. Since D is of principal determinant rank n, these are all non-zero. It follows that F is equivalent in ϕ to a form of type G, and that the *transformation which reduces* $F_1, \cdots, F_n$ *to* $G_1, \cdots, G_n$ *reduces* F *to* G.

The problem now resolves into the simultaneous equivalence in ϕ of $F_1, \cdots, F_n$ to $G_1, \cdots, G_n$.

8. **Transformations which bring sums of pth powers, $p\geqq 3$, into sums of pth powers.** We shall prove the following theorem:

THEOREM XI. *Let ϕ be an (n, p)-proper field. If the non-singular linear transformation $x_i=b_{ij}y_j$ brings a sum of pth powers $a_ix_i^p$, $(a_i\neq0, i=1, \cdots, n; p\geqq3)$, into a sum of pth powers $c_iy_i^p$, $(c_i\neq0; i=1, \cdots, n)$, the matrix $B=(b_{ij})$ of order n has exactly one element in each row and column different from zero.*

Applying the transformation $x_i=b_{ij}y_j$ to $F=a_ix_i^p$ and equating to $c_iy_i^p$ we obtain

$$
a_i(b_{ij}y_j)^p = c_iy_i^p. \tag{20}
$$

Since ϕ is (n, p)-proper we can write (20) in the matrix form

$$
(a_ib_{ij}b_{ik}\cdots b_{im}b_{iq}) = (c_i\delta_{ij}\delta_{ik}\cdots\delta_{iq}), \tag{21}
$$

where $(\delta_{ij}) = (\delta_{ik}) = \cdots = (\delta_{iq}) = I$, and I is a Kronecker delta. Now

$$(a_i b_{i1} b_{ik} \cdots b_{iq}) = (c_i \delta_{i1} \delta_{ik} \cdots \delta_{iq}). \tag{22}$$

Since the element $c_1 \delta_{11} \delta_{11} \cdots \delta_{11}$ is the only non-vanishing element of the matrix on the right of (22), it is of principal determinant rank 1, therefore this is true of the left matrix of (22). We can write the latter matrix as the product

$$(a_i b_{i1} b_{ik} \cdots b_{im})(b_{iq}) \tag{23}$$

of ordinary 2-way matrices, where i is the column index of

$$N = (a_i b_{i1} b_{ik} \cdots b_{im}),$$

and $k \cdots m$ is the partition of indices associated with the rows of N (in (b_{iq}) i and q are row and column indices respectively). Since B is non-singular, (23) implies that N is of rank 1. Every second order minor of N can be written in the form

$$a_r b_{r1} a_s b_{s1} \begin{Vmatrix} b_{rk_1} \cdots b_{rm_1} & b_{rk_2} \cdots b_{rm_2} \\ b_{sk_1} \cdots b_{sm_1} & b_{sk_2} \cdots b_{sm_2} \end{Vmatrix}. \tag{24}$$

By a lemma of another paper* since B is non-singular, the 2-way matrix $Q = (b_{jk} \cdots b_{lm})$ is non-singular, where the partitions $\tau = j \cdots l$ and $\sigma = k \cdots m$ are the partitions of the rows and columns of Q. Since the determinants in (24) for a given r, s, $(r \neq s)$, form the class of all determinant minors of the rows of Q obtained by setting $j = \cdots = l = r, s$ it follows that all of these determinant minors cannot vanish. Hence

$$a_r b_{r1} a_s b_{s1} = 0$$

for $r \neq s$. Since a_r, $a_s \neq 0$, we have $b_{r1} b_{s1} = 0$ for $r \neq s$. Similarly

$$b_{r2} b_{s2} = \cdots = b_{rn} b_{sn} = 0$$

for $r \neq s$. Take $b_{i1} \neq 0$ for i equal to some value t. Then $b_{i1} = 0$ for $i \neq t$. If $b_{i2} \neq 0$ for $i = f$, where $f \neq t$, then $b_{i2} = 0$ for $i \neq f$; similar conclusions hold for $b_{i3}, \cdots, b_{in}$.

It is to be noted that if *p is a prime and ϕ has characteristic p, any non-singular linear transformation brings F into a sum of pth powers.*

Let $B = (b_{ij})$ be chosen so that it satisfies the preceding theorem. Apply the transformation $x_i = b_{ij} y_j$ to a form $c_i x_i^q$. We obtain $c_i (b_{ij} y_j)^q$ which is a sum of qth powers. We have proved the following theorem:

* R. Oldenburger, I, p. 625. See also G. W. King, *The indeterminate and composite products of matrices*, Journal of Mathematics and Physics, vol. 13 (1934), p. 445.

THEOREM XII. *Let ϕ be an (n, p)-proper field. If a non-singular linear transformation in ϕ brings a sum of n pth powers of $x_1, \cdots, x_n$ into such a sum, where $p \geqq 3$, this transformation brings a sum of qth powers of $x_1, \cdots, x_n$, for any q, into a sum of qth powers with the same number of non-vanishing coefficients.*

This implies the following corollary:

COROLLARY I. *Let ϕ be an (n, p)-proper field. Let F be a sum of n pth powers of $x_1, \cdots, x_n$ where $p \geqq 3$. A set of forms $F, R, S, \cdots, T$ in $x_1, \cdots, x_n$ are simultaneously equivalent in ϕ to sums of powers of $x_1, \cdots, x_n$ if and only if $R, S, \cdots, T$ are already sums of powers in $x_1, \cdots, x_n$.*

The reader should compare the simplicity of the theory of equivalence of r-tuples of forms in $x_1, \cdots, x_n$, where one of the forms is equivalent to $a_i x_i^p$, $(a_i \neq 0; i = 1, \cdots, n; p \geqq 3)$, with the corresponding theory for the quadratic case.

We are now able to prove the following theorem:

THEOREM XIII. *Let ϕ be an (n, p)-proper field. If the non-singular transformation $x_i = b_{ij} y_j$, $(i, j = 1, \cdots, n)$, brings a sum of r, $(r \leqq n)$, pth powers $F = a_i x_i^p$, $(i = 1, \cdots, r;\ a_i \neq 0)$, into a sum of r pth powers $F' = b_i y_i^p$, $(i = 1, \cdots, r;\ b_i \neq 0)$, then the matrix $B = (b_{ij})$ is of the form*

$$\left\| \begin{matrix} B_{11} & 0 \\ B_{21} & B_{22} \end{matrix} \right\|,$$

where B_{11} is a minor of order r with exactly one non-vanishing element in each row and column.

The matrix of F is unique since ϕ is also (r, p)-proper. Applying the transformation with matrix B to F we obtain

$$\sum_{i=1}^{r} a_i (b_{ij} y_j)^p, \qquad i = 1, \cdots, n.$$

Equating to F', we can write

$$(25) \qquad \sum_{i=1}^{r} a_i b_{ij} b_{ik} \cdots b_{iq} y_j y_k \cdots y_q = \sum_{i=1}^{r} b_i y_i^p, \qquad j, \cdots, q = 1, \cdots, n.$$

Since there are no terms in $y_{r+1}, \cdots, y_n$ on the right of (25) we obtain, by the restrictions on ϕ,

$$\sum_{i=1}^{r} a_i b_{ij} b_{ik} \cdots b_{iq} = 0,$$

for $j=r+1, \cdots, n$; $k, \cdots, q=1, \cdots, n$. Setting $j=r+1$ in the above equation and writing it in matrix form we get

$$\left(\sum_{i=1}^{r} a_i b_{i,r+1} b_{ik} \cdots b_{im}\right)(b_{iq}) = 0, \tag{26}$$

where $k \cdots m$ is the partition of the rows of the left factor in (26), i is the index of the columns, and i, q are the row and column indices of (b_{iq}). Since B is non-singular the left factor in (26) is of rank zero; whence $a_i b_{i,r+1} b_{ik} \cdots b_{im}=0$ (i not summed) for $i=1, \cdots, r$; $k, \cdots, m=1, \cdots, n$. By an argument used in the proof of Theorem XI, the 2-way matrix $Q=(b_{ik} \cdots b_{lm})$ is non-singular where $\tau=j \cdots l$ is the partition of the rows of Q and $\sigma=k \cdots m$ the partition of the columns. The products $b_{ik} \cdots b_{im}$, i not summed, are the elements in the row of Q obtained by setting $j= \cdots =l=i$. Since Q is non-singular, these do not all vanish for a given i. Hence $a_i b_{i,r+1}=0$ for every i. Similarly $a_i b_{ij}=0$ (i not summed) for $j=r+2, \cdots, n$. Since $a_i \neq 0$ for every i, we have $b_{ij}=0$ for $i=1, \cdots, r$; $j=r+1, \cdots, n$. If B is of this form, the terms on the left of (25) involve only $y_1, \cdots, y_r$. Theorem XI now implies that B_{11} has exactly one element not zero in each row and column.

COROLLARY II. *Let ϕ be an (n, p)-proper field. If the non-singular transformation $x_i=b_{ij}y_j$, $(i, j=1, \cdots, n)$, brings a sum of pth powers $a_i x_i^p$, $(p \geqq 3$; $i=1, \cdots, r$; $a_i \neq 0)$, into a sum of pth powers $b_i y_i^p$, $(i=1, \cdots, r$; $b_i \neq 0)$, then the transformation brings any sum of qth powers $F=c_i x_i^q$, $(i=1, \cdots, r)$, into a sum of qth powers $F'=d_i x_i^q$, $(i=1, \cdots, r)$, where F and F' have the same number of non-vanishing coefficients.*

THEOREM XIV. *Let G be a sum of pth powers, $(p \geqq 3)$, of $x_1, \cdots, x_r$ with non-vanishing coefficients. Let F_i, $(i=1, \cdots, s)$, be a sum of qth powers of $x_1, \cdots, x_r$ for $i=1, \cdots, s$, where $q \geqq 3$ and the coefficients of x_i^p for any i are not all zero in $F_1, \cdots, F_s$. Let ϕ be an (n, p)- and (n, q)-proper field. The class of non-singular transformations $x_i=b_{ij}y_j$, $(i, j=1, \cdots, (n))$, $(n \geqq r)$, which bring G in ϕ into a like sum of powers G', is identical with the class of transformations which bring $F_1, \cdots, F_s$ into a like set $F_1', \cdots, F_s'$.*

Theorem XIII implies that the transformations bringing G into G' bring $F_1, \cdots, F_s$ into $F_1', \cdots, F_s'$. To prove the converse, assume that ϕ contains an infinite number of elements. Write

$$F_i = \mu_{ij} x_j^q, \qquad F_i' = v_{ij} y_j^q, \qquad i = 1, \cdots, s,$$

where by our assumptions $\mu_{1j}, \cdots, \mu_{sj}$ do not all vanish for a given j. Similar properties hold for $V_{1j}, \cdots, V_{sj}$. Write

$$L_j(z) \equiv \sum_{i=1}^{s} \mu_{ij} z_i, \qquad L_j'(z) = \sum_{i=1}^{s} v_{ij} z_i, \qquad j = 1, \cdots, r.$$

Choose a set of values $\alpha_1, \cdots, \alpha_s$ of $z_1, \cdots, z_s$ in ϕ such that* $L_j(\alpha)$, $L_j'(\alpha) \neq 0$, $(j=1, \cdots, r)$. Let $F \equiv \alpha_1 F_1 + \cdots + \alpha_s F_s \equiv L_1(\alpha)x_1^q + \cdots + L_r(\alpha)x_r^q$. The transform $F' = \alpha_1 F_1' + \cdots + \alpha_s F_s'$ of F is $L_1'(\alpha)y_1^q + \cdots + L_r'(\alpha)y_r^q$. Since none of the coefficients of $x_1^q, \cdots, x_r^q, y_1^q, \cdots, y_r^q$ in F and F' vanish, by Theorem XIII, G transforms into a like form G'.

If ϕ is a finite field, embed ϕ in an algebraically closed field ψ containing it. The above argument now applies for ψ. Since the coefficients of G and the b_{ij} are in ϕ, the coefficients of G' are in ϕ.

9. **Equivalence of forms of degree p, $p \geqq 4$, to sums of pth powers.** We shall prove the following theorem:

THEOREM XV. *Let ϕ be an (n, p)-proper field. Let $F = a_{i \cdots tq \cdots m} x_i \cdots x_t x_q \cdots x_m$ be an n-ary form of degree p, $(p \geqq 4)$, for which a subform $F_{\alpha \cdots \xi} = a_{\alpha \cdots \xi q \cdots m} x_q \cdots x_m$ of degree s, $(s \geqq 3)$, is equivalent in ϕ to a sum $S = \mu_i x_i^s$, $(\mu_i \neq 0)$, of n pth powers for some set of values $\alpha, \cdots, \xi$ of $i, \cdots, t$. The form F is equivalent in ϕ to a sum of n sth powers if and only if under reduction of $F_{\alpha \cdots \xi}$ to S in ϕ, F transforms covariantly into a sum of pth powers.*

Let Σ denote the set of values $\alpha, \beta, \cdots, \delta, \rho, \tau, \xi$ of the leading indices $i, \cdots, t$ of F. A study of minors of (a_{iT}), where T is the partition $j \cdots m$, reveals that the principal determinant rank of any subform F_σ (σ here denotes a fixed partition) *is greater than or equal to* the principal determinant rank of a subform F_Σ if σ is contained in Σ. By Theorem I the principal determinant rank of $F_{\alpha \cdots \xi}$ is n. Hence this rank of F is n. By Theorem X, F is equivalent in ϕ to a sum of n pth powers if and only if $F_1, \cdots, F_n$ are simultaneously equivalent in ϕ to sums of pth powers. Since α is contained in the set $\alpha, \cdots, \xi$, the principal determinant rank of F_α is n. Now F_α occurs in the set $F_1, \cdots, F_n$. If F is equivalent in ϕ to a sum of n pth powers, F_α is equivalent in ϕ to a sum of $(p-1)$th powers. There are n such powers because the principal determinant rank of F_α is n. By Theorem X, F_α is equivalent to such a sum if and only if $F_{\alpha 1}, \cdots, F_{\alpha n}$ are simultaneously equivalent in ϕ to sums of $(p-2)$th powers. The subform $F_{\alpha\beta}$ has principal determinant rank n and occurs in this set. Continuing this process, we find that if F is equivalent in ϕ to a sum of pth powers, $F_{\alpha\beta \cdots \tau}$ is equivalent in ϕ to a sum of n $(s+1)$th powers. By Theorem X, this equivalence is valid if and only if $F_{\alpha\beta \cdots \tau 1}, \cdots, F_{\alpha\beta \cdots \tau n}$ are simultaneously equivalent in ϕ to sums of sth powers. The subform $F_{\alpha\beta \cdots \tau\xi}$ occurs in this set.

* O. Haupt, *Einführung in die Algebra*, vol. 1, p. 166.

Reduce $F_{\alpha\beta\ldots\tau\xi}$ non-singularly to a sum of sth powers S. Simultaneously the remaining forms in the set $F_{\alpha\beta\ldots\tau 1}, \cdots, F_{\alpha\beta\ldots\tau n}$ are transformed into a new set $F'_{\alpha\beta\ldots\tau 1}, \cdots, F'_{\alpha\beta\ldots\tau n}$. By the corollary of Theorem XII, these forms are simultaneously equivalent in ϕ to a sum of qth powers if and only if they are already such sums. Assume that they are such sums. By a remark near the end of §7, the transformation T that reduces $F_{\alpha\beta\ldots\tau 1}, \cdots, F_{\alpha\beta\ldots\tau n}$, to sums of sth powers reduces $F_{\alpha\beta\ldots\tau}$ to a sum $F'_{\alpha\beta\ldots\tau}$ of $(s+1)$th powers. Let the result of applying T to $F_{\alpha\beta\ldots\rho 1}, \cdots, F_{\alpha\beta\ldots\rho n}$ be denoted by $F'_{\alpha\beta\ldots\rho 1}, \cdots, F'_{\alpha\beta\ldots\rho n}$. Since $F'_{\alpha\beta\ldots\tau}$ is in this set and is already a sum of n $(s+1)$th powers, the forms $F'_{\alpha\beta\ldots\rho 1}, \cdots, F'_{\alpha\beta\ldots\rho n}$ are, by the corollary of Theorem XII, simultaneously equivalent to sums of $(s+1)$th powers if and only if they are already such sums. If they are such sums, by the remark of §7 referred to above, the transformation T brings $F_{\alpha\beta\ldots\rho}$ into a form $F'_{\alpha\beta\ldots\rho}$, which is a sum of $(s+2)$th powers. Continuing this chain of reasoning we see that the T transforms $F_1', \cdots, F_n'$ of $F_1, \cdots, F_n$ are simultaneously equivlent in ϕ to sums of $(p-1)$th powers if and only if they are already such sums. The transformation T then brings F into a sum of pth powers F'.

Hence if F is equivalent in ϕ to a sum of n pth powers, the transformation T which reduces $F_{\alpha\ldots\xi}$ to a sum of n sth powers must reduce F to a sum of n pth powers. The sufficiency of this condition is obvious.

10. **Simultaneous equivalence of forms of degree p, $(p \geqq 3)$, to sums of pth powers.** The assumption made in Theorem XV concerning the existence of a subform $F_{\alpha\ldots\xi}$ of degree s, $(s \geqq 3)$, equivalent to a sum of n sth powers is evidently not a necessary condition for the equivalence of a form F to a sum of n pth powers. It is therefore convenient to have a method of determining whether or not a set of forms in n variables of degree p, $(p \geqq 3)$, are simultaneously equivalent to sums of pth powers where no form in the set is equivalent to a sum of n pth powers. I have developed such a theory which involves the theorems of §§7–9, Theorem XVI below, and other considerations. Since parts of the method are complicated, the development will be left to the reader. Essential to this theory is the following theorem:

THEOREM XVI. *Let*

$$B = (b_{ij}) = \left\| \begin{matrix} B_{11} & 0 \\ B_{21} & B_{22} \end{matrix} \right\|,$$

where B_{11} is a minor of order r with exactly one non-vanishing element in each row and column. For the form

$$F = a_{ij\ldots km} x_i x_j \cdots x_k x_m, \quad i, \cdots, m = 1, \cdots, n,$$

of degree p, $(p \geqq 3)$, to be equivalent in an (n, p)-proper field ϕ under the non-singular transformation

$$x_i = b_{ij}y_j, \qquad i, j = 1, \cdots, n,$$

with matrix B to a sum of pth powers

$$F' = \lambda_\alpha y_\alpha^p, \qquad \alpha = 1, \cdots, r+g, \qquad \lambda_{r+1}, \cdots, \lambda_{r+g} \neq 0,$$

for some g, it is necessary that the form

$$L = a_{ij\cdots km}x_i x_j \cdots x_k x_m, \qquad i, \cdots, m = r+1, \cdots, n,$$

be equivalent in ϕ under the non-singular transformation

$$x_i = b_{ij}y_j, \qquad i, j = r+1, \cdots, n, \tag{27}$$

with matrix B_{22} to a form of the type

$$L' = \lambda_\alpha y_\alpha^p, \qquad \alpha = r+1, \cdots, r+g.$$

Assume that F' is equivalent in ϕ to F for some choice of the λ's and g under the transformation

$$y_i = b_{ij}x_j, \qquad i, j = 1, \cdots, n, \tag{28}$$

with matrix B. This implies that there exist a transformation (28), λ's, and a g such that

$$\sum_{\alpha=1}^{r+g} \lambda_\alpha b_{\alpha i}b_{\alpha j} \cdots b_{\alpha m}x_i x_j \cdots x_m = F, \qquad i, j, \cdots, m = 1, \cdots, n. \tag{29}$$

Since $b_{\alpha i}=0$ for $\alpha=1, \cdots, r$ and $i=r+1, \cdots, n$, we derive from (29) the relations

$$\sum_{\alpha=r+1}^{r+g} \lambda_\alpha b_{\alpha i}b_{\alpha j} \cdots b_{\alpha m}x_i x_j \cdots x_m = L, \qquad i, \cdots, m = r+1, \cdots, n. \tag{30}$$

If (30) is satisfied, there exists a non-singular transformation

$$y_\alpha = b_{\alpha i}x_i, \qquad \alpha, i = r+1, \cdots, n,$$

and a choice of L' such that L' reduces to L. Hence there exists a non-singular transformation (27) bringing L into L'.

It is to be noted that if $F_1, \cdots, F_t$ are sums of pth powers, then a set of forms $F_1, \cdots, F_t, F_{t+1}, \cdots, F_s$ are simultaneously equivalent in a field ϕ to sums of pth powers if and only if $F_{t+1}, \cdots, F_s$ are equivalent to sums of pth powers under transformations which bring $F_1, \cdots, F_t$ into sums of pth powers. These transformations can be written down by Theorems XIII and XIV.

11. Equivalence of a binary form to a sum of two pth powers. We shall prove a modification of Theorem XV. In this section we shall denote the subform

$$a_{i\,1\cdots1\,e\,h\cdots g\,f\,2\cdots2}\,x_ix_e x_h\cdots x_g x_f, \qquad i,\cdots,f=1,2,$$

of $F=a_{i\cdots m}x_i\cdots x_m$, $(i,\cdots,m=1,2)$, by $F_{\rho\tau}$. In $F_{\rho\tau}$ there are ρ fixed consecutive subscripts equal to 1 and τ equal to 2, the remaining subscripts being free. We shall need the following lemma:

LEMMA III. *Let ϕ be a $(2, p)$-proper field. Let $F=a_{ij\cdots m}x_ix_j\cdots x_m$ be a binary form of degree p with principal determinant rank 2. If the principal determinant rank of a subform $F_{\rho\tau}=x_1F_{\rho+1,\tau}+x_2F_{\rho,\tau+1}$ of degree greater than or equal to 3 is 2, while this rank is 1 for each of the subforms $F_{\rho+1,\tau}$, $F_{\rho,\tau+1}$, the form F is equivalent in ϕ to a sum of pth powers if and only if F is already a sum of pth powers; then $F_{\rho\tau}\equiv F$.*

The columns of the ordinary 2-way matrix $(a_{i\tau})$, where τ is the partition $j\cdots m$, are, by the symmetry of $(a_{i\cdots m})$, the columns of the matrix

$$A'=\left\|\begin{matrix} a_{11\cdots1} & a_{11\cdots12} & a_{11\cdots122} & a_{11\cdots1222}\cdots a_{12\cdots2} \\ a_{21\cdots1} & a_{21\cdots12} & a_{21\cdots122} & a_{21\cdots1222}\cdots a_{22\cdots2}\end{matrix}\right\|;$$

whence the rank of A' is the principal determinant rank of F. Since this rank of F is 2, some minor

$$N=\left\|\begin{matrix} a & b \\ c & d\end{matrix}\right\|=\left\|\begin{matrix} a_{1\,1\cdots1\,2\cdots2\,2\cdots2} & a_{1\,1\cdots1\,1\cdots1\,2\cdots2} \\ a_{2\,1\cdots1\,2\cdots2\,2\cdots2} & a_{2\,1\cdots1\,1\cdots1\,2\cdots2}\end{matrix}\right\|$$

of A' is non-singular, where for each element of the minor matrix N the first subscript is followed by sets of equal numbers, the first containing ρ 1's, the second σ 1's or 2's, and the last τ 2's. N is a minor of the matrix which is given by

$$M=\left\|\begin{matrix} a_{1\,1\cdots1\,u\,v\cdots w\,2\cdots2} \\ a_{2\,1\cdots1\,u\,v\cdots w\,2\cdots2}\end{matrix}\right\|, \qquad u,v,\cdots,w=1,2, \tag{31}$$

where only a typical column of M is displayed in (31). Here the first set of equal subscripts contains ρ 1's and the last τ 2's. Setting $u=1$ and $u=2$ in M we obtain minors M_1 and M_2 which can be written as

$$M_1=\left\|\begin{matrix} a_{1\,1\cdots1\,v\cdots w\,2\cdots2} \\ a_{2\,1\cdots1\,v\cdots w\,2\cdots2}\end{matrix}\right\|, \qquad M_2=\left\|\begin{matrix} a_{1\,1\cdots1\,v\cdots w\,2\cdots2} \\ a_{2\,1\cdots1\,v\cdots w\,2\cdots2}\end{matrix}\right\|.$$

The matrix M is composed of the columns of M_1 and M_2. The first subscript for each element of M_1 is followed by $(\rho+1)$ 1's, and the set of sub-

scripts ends with a set of τ 2's. For the matrix M_2 the sets have ρ 1's and $(\tau+1)$ 2's, respectively. The only columns that M_1 and M_2 do not have in common are the columns of N. Write

$$M_1 = (\xi_1 \cdots \xi_\varrho), \qquad M_2 = (\xi_{\varrho+1} \cdots \xi_{2\varrho}), \qquad N = (\xi_{2\varrho}\xi_1),$$

where the ξ's are columns of M_1, M_2, N. The matrices M, M_1, M_2 are the matrices of the forms $F_{\rho\tau}$, $F_{\rho+1,\tau}$, and $F_{\rho,\tau+1}$ of the lemma. We assume that the principal determinant ranks of $F_{\rho+1,\tau}$, $F_{\rho,\tau+1}$ are 1; whence the ranks of M_1 and M_2 equal 1. Assume that $\xi_2 \neq 0$. Since M_1 is of rank 1, ξ_1, ξ_2 are linearly dependent. Since ξ_2 also occurs among the vectors $\xi_{\varrho+1}, \cdots, \xi_{2\varrho-1}$, and M_2 is of rank 1, ξ_2 and $\xi_{2\varrho}$ are linearly dependent. It follows that ξ_1, $\xi_{2\varrho}$ are linearly dependent contrary to assumption.

Since $\xi_2 = \cdots = \xi_\varrho = 0$, whence $\xi_{\varrho+1} = \cdots = \xi_{2\varrho-1} = 0$, we make use of the assumption of the lemma that the degree of $F_{\rho\tau}$ is greater than or equal to 3, and it therefore follows that there are at least two variables in the set $u, v, \cdots, w$. The elements

$$a_{1\cdots1\,2\cdots2}, \qquad a_{2\,1\cdots1\,2\cdots2},$$

where the set of subscripts of the first element ends with $\tau+1$ 2's and the second element has $\rho+1$ subscripts equal to 1, now occur among the vectors $\xi_2, \cdots, \xi_\varrho$. These elements therefore vanish. Since they are the same as the elements

$$a_{1\,1\cdots1\,2\cdots2\,2\cdots2}, \qquad a_{2\,1\cdots1\,1\cdots1\,2\cdots2}$$

of ξ_1, $\xi_{2\varrho}$, where the subscripts fall into groups as in N, the form $F_{\rho\tau}$ is given by the equation

$$F_{\rho\tau} = a_{1\,1\cdots1\,1\cdots1\,2\cdots2}\, x_1^{\sigma+1} + a_{2\,1\cdots1\,2\cdots2\,2\cdots2}\, x_2^{\sigma+1}. \tag{32}$$

Since $F_{\rho\tau}$ is thus a sum of $(\sigma+1)$th powers, $(\sigma+1 \geqq 3)$, by Theorem XV, F is equivalent to a sum of pth powers if and only if it is already such a sum.

To complete the proof of the lemma we note that if F is a sum of two pth powers, the principal determinant rank of every subform $F_{\rho\tau} \not\equiv F$ is 1 or 0; whence the form $F_{\rho\tau}$ of the lemma is F.

The *characteristic determinant* of two quadratic forms F, G in the same variables x_1, x_2 with matrices A, B, respectively, is the determinant $|x_1A + x_2B|$.

If F is equivalent to a sum of 2 pth powers and there is no subform $F_{\rho\tau}$ satisfying the conditions of Lemma III, there is a quadratic form $F_{\rho\tau}$ with rank 2. For a form F having such a subform we shall prove the following strengthened form of Theorems III and VI. It is to be noted that *no distinction is made between even and odd values of p.*

THEOREM XVII. *Let ϕ be a $(2, p)$-proper field. Let $F = a_{ij\cdots km}x_i x_j \cdots x_k x_m$, $(i, \cdots, m = 1, 2)$, be a binary form of degree p, $(p \geqq 3)$, not already a sum of pth powers. The form F is equivalent in ϕ to a sum of 2 pth powers if and only if for some ρ, τ the quadratic subform $F_{\rho\tau}$ is of rank 2, the characteristic determinant D of $F_{\rho\tau}$, $F_{\rho-1,\tau+1}$ or $F_{\rho\tau}$, $F_{\rho+1,\tau-1}$ factors in ϕ into distinct linear factors, and under reduction of D to canonical form $Kx_1' x_2'$ in ϕ F transforms covariantly into a sum of pth powers.*

That the rank of $F_{\rho\tau}$ equals 2 for some ρ, τ was noted above to be a necessary condition. By the argument used in the proof of Theorem XV the simultaneous equivalence in ϕ of $F_{\rho\tau}$, $F_{\rho-1,\tau+1}$ and $F_{\rho\tau}$, $F_{\rho+1,\tau-1}$ to sums of squares is also a necessary condition. By Theorem X these pairs of forms are simultaneously equvalent in ϕ to sums of squares if and only if the cubic forms $F_{\rho-1,\tau} = x_1 F_{\rho\tau} + x_2 F_{\rho-1,\tau+1}$ and $F_{\rho,\tau-1} = x_1 F_{\rho+1,\tau-1} + x_2 F_{\rho\tau}$ are respectively equivalent to sums of cubes. By Theorem IIIa this equivalence is possible if and only if the characteristic determinants of $F_{\rho\tau}$, $F_{\rho-1,\tau+1}$ and $F_{\rho\tau}$, $F_{\rho+1,\tau-1}$, respectively, factor in ϕ into distinct linear factors. If it is assumed that this necessary condition is satisfied for one of the pairs of quadratic subforms, one of the cubic forms $F_{\rho-1,\tau}$ or $F_{\rho,\tau-1}$ is equivalent in ϕ to a sum of 2 cubes. By Theorem III the transformation in ϕ, which reduces D to canonical form $Kx_1' x_2'$, simultaneously reduces $F_{\rho-1,\tau}$ or $F_{\rho,\tau-1}$, as the case may be, to sums of cubes. By Theorem XV, F is now equivalent in ϕ to a sum of pth powers if and only if under reduction of D to $Kx_1' x_2'$. F transforms into a sum of pth powers.

ARMOUR INSTITUTE OF TECHNOLOGY,
CHICAGO, ILL.
THE INSTITUTE FOR ADVANCED STUDY,
PRINCETON, N. J.

TWO-DIMENSIONAL SPACES IN WHICH THERE EXIST CONTIGUOUS POINTS*

BY

E. C. KLIPPLE

In a recent number of The Rice Institute Pamphlet† R. L. Moore has formulated a set of axioms in terms of the undefined notions "point," "region," and "contiguous to." These axioms‡ (Axioms A, B, C, 0, 1, and 2 of this paper) serve as the basis for the proofs of a considerable number of theorems of ordinary point set theory, including a large proportion of the theorems of the first two chapters of Moore's book.§ Nevertheless, there exist spaces satisfying these axioms in which an arc may contain only a finite number of points and in which a region may consist of a finite number of points.

In the present paper a study is made of spaces which satisfy the above mentioned axioms and some additional axioms which restrict the spaces to being, in a certain sense, two dimensional. The ordinary euclidean plane is a space which satisfies all the axioms.

I wish to acknowledge my indebtedness to Professor R. L. Moore, and to thank him for suggesting the problem and for many helpful criticisms in the course of its development.

For definitions of terms used but not defined here the reader may refer to S.C.P.

DEFINITIONS. *A simple closed curve is a compact continuum, containing at least two distinct non-contiguous points, which is disconnected by the omission of any two of its non-contiguous points. A triune is a set of three distinct points such that each of them is contiguous to each of the others.*

AXIOM A. *No point is contiguous to itself.*

AXIOM B. *If the point A is contiguous to the point B, then B is contiguous to A.*

---

* Presented to the Society, October 28, 1933; received by the editors August 18, 1937.

† Vol. 23 (1936), no. 1. In the present treatment the abbreviation S.C.P. will be used to designate part 1 of this paper.

‡ It being understood that in the statement of Axiom 2 of S.C.P. the word "non-degenerate" is to be omitted. It is clear from the context that the retention of this word was not intended.

§ *Foundations of Point Set Theory*, American Mathematical Society Colloquium Publications, vol. 13, New York, 1932. In the present treatment the abbreviation P.S.T. will be used to designate this book.

AXIOM C. *If M is a closed point set and every point of the set H is contiguous to some point of M, then no point of $S-M$ is a limit point of H.*

AXIOM 0. *Every region is a point set.*

AXIOM 1. *There exists a sequence $G_1, G_2, G_3, \cdots$ such that (1) for each n, G_n is a collection of regions covering S, (2) for each n, G_{n+1} is a subcollection of G_n, (3) if R is any region whatsoever, and if X is a point of R and Y is a point of R either identical with X or not, then there exists a natural number m such that if g is any region belonging to the collection G_m and containing X, then $\bar{g}$ is a subset of $(R-Y)+X$, (4) if $M_1, M_2, M_3, \cdots$ is a sequence of closed point sets such that M_n contains M_{n+1} for each n and there exists a region g_n of the collection G_n such that M_n is a subset of $\bar{g}_n$ for each n, then there is at least one point common to all the point sets of the sequence $M_1, M_2, M_3, \cdots$.*

AXIOM 2. *If P is a point of a region R, there exists a connected domain containing P and lying in R.*

AXIOM 3. *If J is a simple closed curve or triune, then $S-J$ is the sum of two mutually separated connected point sets such that J is the boundary of each of them.*

In this paper I deal very frequently with complementary domains of simple closed curves and triunes. If J is a simple closed curve or triune and ω is a point of $S-J$, then by *the interior of J with respect to ω*, is meant that complementary domain of J which does not contain ω. Similarly, that complementary domain of J which contains ω is called *the exterior of J with respect to ω*. In case no ambiguity arises, the terms *interior* and *exterior* of J will be used without making specific reference to a point ω.

THEOREM 1. *Let J_1 and J_2 denote two point sets each of which is either a simple closed curve or a triune. Let I_1 and I_2 denote the interiors of J_1 and J_2, respectively, and suppose J_2 is a subset of J_1+I_1. Then I_2 is a subset of I_1.*

THEOREM 2. *If J is a simple closed curve, and AB separate C and D on J, and AXB is an arc such that the segment AXB is a subset of I, the interior of J, and no point of the segment AXB is constiguous to any point of $J-(A+B)$, then (1) I_1, the interior of $AXBCA$, is a subset of I, (2) the segment ADB is a subset of the exterior of $AXBCA$, and (3) I_1 has no point in common with I_2, the interior of $AXBDA$.*

THEOREM 3. *Under the hypothesis of Theorem 2, $I=I_1+I_2+$segment AXB.*

Proof. Suppose $I(I_1+I_2+\text{segment } AXB)=M$, where M is a non-vacuous point set. Let P be a point of M. By Theorem 38 of S.C.P. there exists an arc PX from P to X lying in I. The arc PX contains an arc PX' such that

$PX'-X'$ is a subset of M, and X' is a point of the segment AXB. Similarly, there exists an arc $P\omega$ lying in E_2, the exterior of $AXBDA$. The arc $P\omega$ contains an arc PC', such that $PC'-C'$ is a subset of M, and C' is a point of the segment ACB. Let T denote the first point in the order from X' to P which $X'P$ has in common with PC'. If no point of the interval $X'T$ of $X'P$, except possibly T, is contiguous to any point of $(TC'-T)$, where TC' denotes the interval of PC' from T to C', then $X'T+TC'$ is an arc from X' to C'. If there exist points of $X'T-T$ which are contiguous to points of $TC'-T$, there must be a first such point in the order from X' to T. For otherwise, there would be infinitely many such points, and the set of all such points would have a limit point in $X'T-T$. But by Axiom C, every limit point of such a set must belong to TC', and a contradiction is reached. Let W denote the first point of $X'T-T$ which is contiguous to a point of $TC'-T$. By Axiom C, W is contiguous to only a finite number of points of $TC'-T$. Let V denote the last point in the order from T to C' which is contiguous to W. Let $X'W$ denote the interval of $X'T$ from X' to W or the point X' according as W is not or is identical with X'. Let VC' denote the interval of TC' from V to C' or the point C' according as V is not or is identical with C'. The point set $X'W+VC'$ is an arc containing at least three points. Thus, in any case, the point set $PX'+PC'$ contains an arc $X'P'C'$, such that the segment $X'P'C'$ contains at least one point and is a subset of M. Similarly, we may show the existence of an arc $X''P''C''$ such that (1) the segment $X''P''C''$ contains at least one point and is a subset of I_1 and (2) the points X'' and C'' are points of the segments AXB and ACB, respectively. Let $X'X''$ denote the point X' or the arc of the segment AXB from X' to X'', according as X'' is or is not identical with X'. Also let $C'C''$ denote the point C' or the arc of the segment ACB from C' to C'', according as C'' is or is not identical with C'. By means of repeated applications of Axiom C it may be shown that the point set $X'P'C'+C'C''+X''P''C''+X'X''$ contains a simple closed curve J', which contains at least one point of each of the segments $X'P'C'$ and $X''P''C''$. Let I' denote the interior of J'. Now I' is a subset of I by Theorem 1. Thus the segment ADB is a subset of E', the exterior of J'. Since the connected set I_2 plus the segment ADB contains no point of J' but does contain a point of E', it follows that I_2 plus the segment ADB is a subset of E'. Furthermore I' cannot contain a point of the segment AXB. For, suppose I' contains the point F of the segment AXB. The connected set $F+I_2$ contains no point of J' but contains points of both complementary domains of J'. We conclude from this contradiction that the segment AXB is a subset of $J'+E'$. Now I' cannot be a subset of M since there exists a point of $J'\cdot I_1$ which is either a limit point of I' or contiguous to a point of I',

and M and I_1 are two mutually separated domains. Also, I' cannot be a subset of I_1 since there exists a point of $J' \cdot M$ which is either a limit point of I' or contiguous to a point of I', and M and I_1 are two mutually separated domains. Also I' cannot be a subset of $M+I_1$ and contain points of both M and I_1 since I' would thus be the sum of two mutually separated sets, contrary to Axiom 3. Thus in any case we reached a contradiction, and the theorem is established.

THEOREM 4. *If the points A and B separate the points C and D on the simple closed curve J, and if the segments AXB and CYD are both subsets of I, the interior of J, then these segments have at least one point in common.*

Proof. Suppose the segments AXB and CYD have no point in common. There exists an arc $A'\omega$ such that $A'\omega - A'$ is a subset of E, the exterior of J, and such that A' is a point of the segment CAD of J. Similarly, there exists an arc $B'\omega$ such that $B'\omega - B'$ is a subset of E and B' is a point of the segment CBD. The point set $A'\omega + B'\omega$ contains an arc $A'X'B'$ such that the segment $A'X'B'$ is a subset of E. Let AA' denote the point A or the arc of the segment CAD from A to A' according as A' is or is not A, and let BB' denote the point B or the arc of the segment CBD from B to B' according as B' is or is not B. It may be shown that there exists a simple closed curve J^* satisfying the following conditions: (1) J^* is a subset of $AA'+BB'+AXB+A'X'B'$, (2) J^* contains at least one point of each of the segments AXB and $A'X'B'$, and (3) $J^* \cdot J$ is the sum of two mutually separated continua which separate C and D on J; therefore $J - J^* \cdot J = g_1 + g_2$, where g_1 and g_2 are mutually separated segments containing C and D, respectively. Since J^* contains no point of the arc CYD, it follows that the connected point set $CYD+g_1+g_2$ is a subset of I^*, a complementary domain of J^*. Thus J is a subset of I^*+J^*. Hence, by Theorem 1, either I or E is a subset of I^*. But each of the sets I and E contains a point of J^*, and J^* contains no point of I^*. Thus we reach a contradiction, and the theorem is established.

THEOREM 5. *If J is a simple closed curve or triune, then I, the interior of J, contains infinitely many points.*

Proof. Suppose I contains exactly n points, where n denotes a natural number. Thus every point of J is contiguous to some point of I. If there are infinitely many points of J, the closed point set I contains a limit point of J by Axiom C. But this is impossible since J is closed. Hence J contains only a finite number of points. Let A and B be two contiguous points of J, and let X_0 be a point of $J-(A+B)$. Let X be a point of I which is contiguous to A and Y a point of I contiguous to B. Let XY denote the point X or an arc from X

to Y lying in I, according as Y is or is not identical with X. The point set $XY+A+B$ contains a simple closed curve or triune J_1 which is a subset of $J+I$ and contains A, B, and a point X_1 of I but is such that X_0 is in the exterior of J_1. Similarly, there exists a simple closed curve or triune J_2 which is a subset of J_1 plus its interior and contains A, B, and a point X_2 of the interior of J_1 but is such that X_1 is in the exterior of J_2. By continuing the indicated process $n+1$ times we reach a contradiction, since $X_1, X_2, X_3, \cdots, X_{n-1}$ are distinct points of I. Hence I must contain infinitely many points.

THEOREM 6. *Let J denote a simple closed curve whose interior I contains a point P which is contiguous to at least three distinct points of J. Let $P_1, P_2, \cdots, P_n$, $(n \geqq 3)$, be points of J (in the order indicated if $n>3$), and let $\beta = P_1+P_2+\cdots+P_n$. Suppose β is the set of all points of J that are contiguous to P. Let P_kP_{k+1}, $(k=1, 2, \cdots, n)$, denote that arc of J which contains only the points P_k and P_{k+1} of the set β, and let P_{n+1} denote P_1. Let I_k, $(k=1, 2, \cdots, n)$, denote the interior of the triune or simple closed curve $P+P_kP_{k+1}$. Then $I = P+\sum_{k=1}^{n} I_k$.*

Proof. By Theorem 1, I_k, $(k=1, 2, \cdots, n)$, is a subset of I. Also, if $k \neq j$, then I_k and I_j are mutually exclusive. Suppose $I-(P+\sum_{k=1}^{n} I_k) = M$, where M is a non-vacuous point set, and let M_1 denote a component of M. Now M is a domain; hence M_1 is a domain. Let α denote the set of all points of J, each of which is either a limit point of M_1 or contiguous to a point of M_1. There exists at least one point of α. For let Q_1 denote any point of M_1, and let $Q_1\omega$ denote an arc from Q_1 to ω lying in the exterior of the simple closed curve or triune $P+P_1P_2$. Let Q_1' denote the first point that $Q_1\omega$ has, in order Q_1 to ω, in common with J. Now Q_1' is obviously a point of α, since Q_1' is the first point of $Q_1\omega$, in order Q_1 to ω, which does not belong to M_1.

Suppose there exist two points X and Y of α, which do not lie together on one of the arcs $P_1P_2, P_2P_3, \cdots, P_nP_1$. Since X and Y are non-contiguous points of J, it follows that J is the sum of two arcs from X to Y having nothing in common except their end points such that the corresponding segments of these arcs are mutually separated. There exists a point P_r of β on one of these segments and a point P_s of β on the other. Since P_r and P_s are both points of β, P_r+P+P_s is an arc lying in $I+J-(M+X+Y)$. There exists an arc P_rQP_s which is a subset of $J+E-(X+Y)$, where E denotes the exterior of J, and is such that $P_rQP_s \cdot J$ is the sum of two mutually separated connected point sets M_r and M which contain P_r and P_s, respectively, but no other points of β, and which separate X and Y on J. The point set $P+P_rQP_s$ is a simple closed curve J' which contains P and is such that $J' \cdot J$ is the sum of two mutually separated connected point sets which separate X and Y on J.

Thus $J-J'\cdot J$ is the sum of two mutually separated segments g_X and g_Y which contain X and Y, respectively. Now J' separates X from Y, for otherwise g_X and g_Y would lie together in I', a complementary domain of J'; therefore by Theorem 1, either I or E would be a subset of I', which is contrary to the fact that both I and E contain points of J'. Since X and Y are boundary points of M_1 lying in different complementary domains of J', it follows that there are points of M_1 lying in different complementary domains of J'. This is impossible since M_1 contains no point of J'. From this contradiction we conclude that if X and Y are any two distinct points of α, then there exists an arc of the set $P_1P_2, P_2P_3, \cdots, P_nP_1$ which contains both X and Y.

Thus, if $J-\beta$ contains a point X of α, then the arc of the set $P_1P_2, P_2P_3, \cdots, P_nP_1$ which contains X must contain all of α. Now if α is not a subset of one of the arcs of the set $P_1P_2, P_2P_3, \cdots, P_nP_1$, it follows that $J-\beta$ contains no point of α. Furthermore, it follows that $n=3$ and $\alpha=\beta=P_1+P_2+P_3$. In this case, since J is not a triune, one of the segments P_1P_2, P_2P_3, P_3P_1 exists. Suppose the segment P_3P_1 is non-vacuous. Using methods similar to those used in the early part of the proof to get J', we obtain a simple closed curve J^* satisfying the following conditions: (1) J^* is a subset of $P+J+I_3+E$, (2) J^* contains P, P_2, a point of the segment P_3P_1, a point of E, and a point of I_3, and (3) $J^*\cdot J$ is the sum of two mutually separated connected point sets which separate P_1 and P_3 on J, and $J-J^*\cdot J$ is therefore the sum of two mutually separated segments, one containing P_1 and the other containing P_3. The argument used above to show that J' separates X and Y may be applied here to show that J^* separates P_1 and P_3. Hence J^* separates two points P_1' and P_3' of M_1, since P_1 and P_3 are points of α. But this is impossible since M_1 is connected and contains no point of J^*. Thus we reach a contradiction, and we conclude that there exists an arc P_iP_{i+1} of the set of arcs $P_1P_2, P_2P_3, \cdots, P_nP_1$ which contains α.

Let E_i denote the exterior of the simple closed curve or triune $(P+P_iP_{i+1})$. The domain E_i contains M_1, and we write $E_i=M_1+(E_i-M_1)$. The boundary of M_1 is a subset of $(P+P_iP_{i+1})$. Hence M_1 and E_i-M_1 are mutually separated, contrary to Axiom 3. Thus we reach a contradiction, and the theorem is established.

THEOREM 7. *Let $P_1, P_2, \cdots, P_n$, $(n\geqq 3)$, be points of the simple closed curve J (in the order indicated if $n>3$). Let $A_1, A_2, \cdots, A_k$, $(k\geqq 2)$, be points of the arc A_1A_k (in the order indicated if $k>2$) where A_1A_k is a subset of I, the interior of J. Let $\beta=P_1+P_2+\cdots+P_n$, and let $\gamma=A_1+A_2+\cdots+A_k$. Suppose β is the set of all points of J each of which is contiguous to at least one point of A_1A_k, and suppose γ is the set of all points of A_1A_k each of which is contiguous to at least one point of β. Suppose furthermore that no point of β is contiguous to both*

A_1 and A_k, but that P_1 is contiguous to A_1 and to no other point of γ and P_n is contiguous to A_k and to no other point of γ. It follows that if $j<j'$, and if P_j is contiguous to A_i and $P_{j'}$ is contiguous to $A_{i'}$, then $i \leqq i'$. Let P_jP_{j+1}, $(j=1, 2, \cdots, n)$, denote that arc of J from P_j to P_{j+1} which contains no other point of β, where P_{n+1} denotes P_1. Let A_iA_{i+1}, $(i=1, 2, \cdots, k-1)$, denote the arc of A_1A_k from A_i to A_{i+1}. Let I_j, $(j=1, 2, \cdots, n-1)$, denote the interior of C_j, where C_j denotes the triune or simple closed curve $(P_jP_{j+1}+A_i)$ in case P_j and P_{j+1} are both contiguous to A_i, or C_j denotes the simple closed curve $(P_jP_{j+1}+A_iA_{i+1})$ in case no point of γ is contiguous to both P_j and P_{j+1}, but where P_j is contiguous to A_i and P_{j+1} is contiguous to A_{i+1}. Let I_n denote the interior of the simple closed curve $(P_nP_1+A_1A_k)$. If P_j is contiguous to each of the points A_{ij}, $A_{ij+1}, \cdots, A_{ij+kj}$, let $I_{j1}, I_{j2}, \cdots, I_{jkj}$ denote the interiors of $(P_j+A_{ij}A_{ij+1})$, $(P_j+A_{ij+1}A_{ij+2}), \cdots$, and $(P_j+A_{ij+kj-1}A_{ij+kj})$, respectively. Then

$$I = A_1A_k + \sum_{j=1}^{n} I_j + \sum_{j=1}^{n} \sum_{t=1}^{kj} I_{jt},$$

where I_{jt} is a null set if P_j is contiguous to only one point of γ.

The theorem is illustrated by Fig. 1.

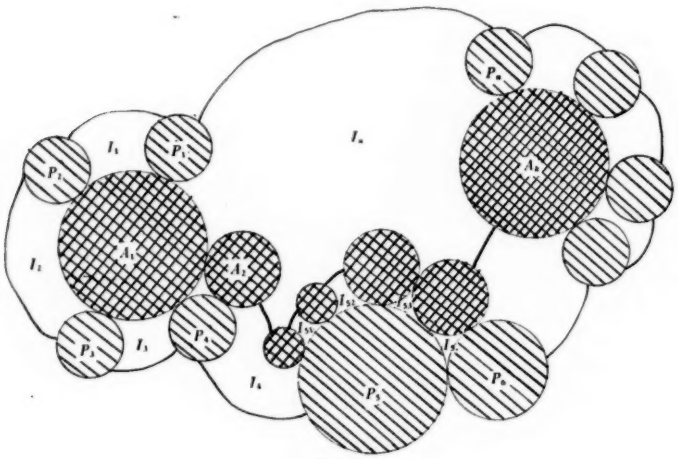


FIG. 1

Proof. First I shall prove the assertion made in the statement of the theorem to the effect that if $j<j'$, and if P_j is contiguous to A_i and $P_{j'}$ is contiguous to $A_{i'}$, then $i \leqq i'$. Suppose there exist two integers j and j' such that

$j<j'$, P_j and $P_{j'}$ are contiguous to A_i and $A_{i'}$, respectively, and $i>i'$. It follows that $i>1$ and $i'<k$. Also since P_1 is contiguous to A_1 but to no other point of γ, we see that $1<j$. Again, since P_n is contiguous to A_k but to no other point of γ, we see that $j'<n$. Thus $1<j<j'<n$; hence P_1 and $P_{j'}$ separate P_j and P_n on J. Let $A_{i_1'}$ be the point of γ with lowest subscript which is contiguous to $P_{j'}$, and let A_{i_1} be the point of γ with the highest subscript which is contiguous to P_j. Let $A_1A_{i_1'}$ denote the arc $A_1A_{i_1'}$ of A_1A_k or the point A_1 according as i_1' is not or is 1, and let $A_{i_1}A_k$ denote the arc $A_{i_1}A_k$ of A_1A_k or the point A_k according as i_1 is not or is k. Now $P_1+A_1A_{i_1'}+P_{j'}$ and $P_j+A_{i_1}A_k+P_n$ are arcs satisfying the conditions of Theorem 4. Hence $A_1A_{i_1'}$ and $A_{i_1}A_k$ have a point in common. But $1\leqq i_1'\leqq i'<i\leqq i_1\leqq k$; therefore $A_1A_{i_1'}$ and $A_{i_1}A_k$ cannot have a point in common. Thus we reach a contradiction, and we conclude that if $j<j'$, and if P_j and $P_{j'}$ are contiguous to A_i and $A_{i'}$, respectively, then $i\leqq i'$.

This result enables us to verify the following implications made in the statement of the theorem: (1) there do not exist two distinct points of γ each of which is contiguous to both P_j and P_{j+1}; (2) if no point of γ is contiguous to both P_j and P_{j+1}, then there exists an integer i such that A_i is contiguous to P_j and A_{i+1} is contiguous to P_{j+1}; and (3) if no point of β is contiguous to both A_i and A_{i+1}, then there exists an integer j such that $j<n$ and P_j and P_{j+1} are contiguous to A_i and A_{i+1}, respectively.

Suppose $I-(A_1A_k+\sum_{j=1}^{n}I_j+\sum_{j=1}^{n}\sum_{t=1}^{k_j}I_{jt})=M$ is a non-vacuous point set. Let M_1 be a component of the domain M, and let α denote the boundary of M_1. Obviously α is a subset of $J+A_1A_k$. Furthermore, α contains at least one point of J and at least one point of A_1A_k. For let P be a point of M_1, and let $P\omega$ be an arc from P to ω lying in the exterior of $(A_1A_k+P_nP_1)$. Now $P\omega$ obviously contains a point of $\alpha\cdot J$. Similarly, let PA_1 denote an arc from P to A_1 which lies in I. Now PA_1 contains a point of $\alpha\cdot A_1A_k$.

Suppose J contains two non-contiguous points X and Y of α. If X and Y do not lie together on any one of the arcs P_1P_2, P_2P_3, $\cdots$, P_nP_1, there exist two points P_r and P_s of β which separate X and Y on J; hence there exists a simple closed curve J' such that: (1) J' is a subset of $E+J+A_1A_k$; (2) $J'\cdot J=M_r+M_s$ where M_r and M_s are two mutually separated continua which contain P_r and P_s, respectively, and which separate X and Y on J; therefore $J-J'\cdot J=g_X+g_Y$ where g_X and g_Y are mutually separated segments containing X and Y, respectively; (3) J' contains at least one point O of E and at least one point of A_1A_k. If X and Y lie together on the arc P_jP_{j+1} of the set P_1P_2, P_2P_3, $\cdots$, P_nP_1, there exists a simple closed curve J' such that: (1) J' is a subset of E plus the segment XY of $P_jP_{j+1}+I_j+A_1A_k+(J-P_jP_{j+1})$; (2) $J'\cdot J$ is the sum of two mutually separated continua which separate X

and Y on J; hence $J-J'\cdot J=g_X+g_Y$ where g_X and g_Y are mutually separated segments containing X and Y, respectively; (3) J' contains at least one point O of E and at least one point of A_1A_k. Now, in either case it may be readily shown that J' separates X from Y and hence that J' separates two points X' and Y' of M_1. But this is impossible since M_1 is connected and contains no point of J'. From this contradiction we conclude that $\alpha\cdot J$ consists either of a single point or of two contiguous points. By an analogous argument it may be shown that $\alpha\cdot A_1A_k$ consists either of a single point or of two contiguous points.

Now I wish to show that α is a subset of a simple closed curve or triune J^*, which is a subset of $J+A_1A_k$, satisfying the condition that either there exists an integer j such that I_j is the interior of J^*, or there exists a pair of integers (j, t) such that I_{jt} is the interior of J^*. If $\alpha\cdot J$ is a subset of P_nP_1, then $J^*=P_nP_1+A_1A_k$. If $\alpha\cdot J$ is not a subset of P_nP_1, we shall use the following procedure.

Case 1. Suppose that P_jP_{j+1}, $(j\neq n)$, is the only arc of the set P_1P_2, P_2P_3, $\cdots$, P_nP_1 which contains $\alpha\cdot J$. It follows that if X denotes either P_j or P_{j+1}, then $P_jP_{j+1}-X$ contains at least one point Z of α. If P_j and P_{j+1} are both contiguous to the point A_i of γ, then we may show that $A_i=\alpha\cdot A_1A_k$. Suppose the contrary. Let Y denote a point of $\alpha\cdot A_1A_k$ which is different from A_i. Now Y either precedes or follows A_i, in the order A_1 to A_k. Suppose Y precedes A_i. Then $i\neq 1$ and $j\neq 1$, and Y is not contiguous to P_{j+1}. There exists a simple closed curve J' having the following properties: (1) J' is a subset of $J+A_1A_k+I_n$; (2) $J'\cdot J$ is a connected point set such that the connected point set $J-J'\cdot J$ contains $P_n+(P_jP_{j+1}-P_j)$; (3) $J'\cdot A_1A_k$ is a connected set such that $A_1A_k-J'\cdot A_1A_k$ is either a connected set containing Y and A_1, or the sum of two mutually separated connected sets, one containing Y and A_1 and the other containing A_k; (4) J' contains at least one point of I_n and at least one point of $J-P_nP_1$. By means of these four conditions imposed on J' we may readily show that J' separates Y from Z and hence that J' separates two points Y' and Z' of M_1. But this is impossible since M_1 is connected and contains no point of J'. Thus we conclude that Y cannot precede A_i on A_1A_k. Similarly Y cannot follow A_i on A_1A_k. Therefore $A_i=\alpha\cdot A_1A_k$. Thus $(A_i+P_jP_{j+1})$ is a simple closed curve or triune J^* which contains α and whose interior is I_j.

If there exists no point of γ which is contiguous to both P_j and P_{j+1}, then there exists an integer i such that P_j is contiguous to A_i and P_{j+1} is contiguous to A_{i+1}. In this case we show that $\alpha\cdot A_1A_k$ is a subset of A_iA_{i+1}. Suppose the contrary. Let Y denote a point of $\alpha\cdot A_1A_k$ not on A_iA_{i+1}. Either Y precedes A_i or Y follows A_{i+1}, in the order A_1 to A_k. Suppose Y precedes A_i.

Then $i \neq 1$ and $j \neq 1$ or n. Now there exists a simple closed curve J' having the following properties: (1) J' is a subset of $J+A_1A_k+I_n$; (2) $J' \cdot J$ is a connected point set such that the connected set $J-J' \cdot J$ contains $P_n+(P_jP_{j+1}-P_j)$; (3) $J' \cdot A_1A_k$ is connected, and $A_1A_k-J' \cdot A_1A_k$ is the sum of two mutually separated connected point sets, one containing Y and A_1, and the other containing A_{i+1} and A_k; (4) J' contains at least one point of I_n and at least one point of $J-P_nP_1$. Thus again J' separates Y from Z and hence separates two points Y' and Z' of M_1. Again we reach a contradiction, and we conclude that Y cannot precede A_i. The same argument applies to show that Y cannot follow A_{i+1}. Hence A_1A_k is a subset of A_iA_{i+1}, and $(P_jP_{j+1}+A_iA_{i+1})$ is therefore a simple closed curve J^* containing α. The interior of J^* is I_j.

Case 2. If there exists no integer r such that P_rP_{r+1} is the only arc of the set $P_1P_2, P_2P_3, \cdots, P_nP_1$ which contains $\alpha \cdot J$, it follows that $\alpha \cdot J$ is a point P_j of β. The case in which $\alpha \cdot J = P_n$ has been disposed of in the paragraph preceding Case 1. Hence suppose $j < n$. Let A_i be the point of γ with lowest subscript which is contiguous to P_j, and let $A_{i'}$ be the point of γ with highest subscript which is contiguous to P_j. By means of an argument like that used in Case 1, it may be shown that no point of $\alpha \cdot A_1A_k$ precedes A_i or follows $A_{i'}$ in the order from A_1 to A_k. If $i = i'$, then $\alpha \cdot A_1A_k = A_i$. Thus $(P_jP_{j+1}+A_i)$ or $(P_jP_{j+1}+A_iA_{i+1})$, according as P_j and P_{j+1} are or are not both contiguous to A_i, is a triune or simple closed curve J^* which contains α and whose interior is I_j. If $i \neq i'$, then the arc $A_iA_{i'}$ of A_1A_k contains $\alpha \cdot A_1A_k$. Furthermore, since $\alpha \cdot A_1A_k$ consists either of a single point or of two contiguous points, it follows that $\alpha \cdot A_1A_k$ is a subset of one of the arcs of the set $A_iA_{i+1}, A_{i+1}A_{i+2}, \cdots, A_{i'-1}A_{i'}$. Suppose the arc $A_{i+t-1}A_{i+t}$ contains $\alpha \cdot A_1A_k$. Then since A_s, $(i \leqq s \leqq i')$, is contiguous to P_j, it follows that $(P_j+A_{i+t-1}A_{i+t})$ is a simple closed curve or triune J^* which contains α and whose interior is I_{jt}. Thus for any case it has been shown that α is a subset of a simple closed curve or triune J^* satisfying the condition that either there exists an integer j such that I_j is the interior of J^*, or there exists a pair of integers (j, t) such that I_{jt} is the interior of J^*.

Let E^* denote the exterior of J^*. Now E^* contains M_1; hence $E^* = M_1 + (E^* - M_1)$, where both M_1 and $E^* - M_1$ are non-vacuous point sets neither of which contains a point of α. Thus E^* is the sum of two mutually separated sets, contrary to Axiom 3; and the theorem is established.

THEOREM 8. *Let the following changes, and none other, be made in the hypotheses of Theorem* 7: $A_1, A_2, \cdots, A_k$, $(k \geqq 1)$, *are points of an arc* A_1T *(in the order indicated if* $k > 1$*) where* $A_1T - T$ *is a subset of* I *and* T *is a point of*

the segment P_nP_1, $(n \geqq 2)$. The set $\gamma = A_1 + A_2 + \cdots + A_k$ is the set of all points of $A_1T - T$ which are contiguous to points of J; $\beta = P_1 + P_2 + \cdots + P_n$ is the set of all points of J which are contiguous to points of γ. The point P_n is no longer restricted to be contiguous to A_k alone of the set γ; I_n denotes the interior of $P_nT + A_kT$; and I_0 denotes the interior of $A_1T + TP_1$. Then

$$I = (A_1T - T) + \sum_{j=0}^{n} I_j + \sum_{j=1}^{n} \sum_{t=1}^{kj} I_{jt}.$$

If $\beta = P_1 + P_2 + P_3 + \cdots$ and $\gamma = A_1 + A_2 + A_3 + \cdots$, where β and γ are infinite sets each having T as its only limit point, then

$$I = (A_1T - T) + \sum_{j=0}^{\infty} I_j + \sum_{j=1}^{\infty} \sum_{t=1}^{kj} I_{jt}.$$

THEOREM 9. *Let the following changes, and none other, be made in the hypotheses of Theorem 7: V, A_1, A_2, $\cdots$, A_k, T, $(k \geqq 1)$, are points of the arc VA_1T in the order indicated, where the segment VA_1T is a subset of I, and where V and T are points of J in the order V, P_1, P_2, $\cdots$, P_n, T, $(n \geqq 1)$. The set $\gamma = A_1 + A_2 + \cdots + A_k$ is the set of all points of the segment VA_1T which are contiguous to points of J, and $\beta = P_1 + P_2 + \cdots + P_n$ is the set of all points of J which are contiguous to points of γ. The points P_1 and P_n are no longer restricted as to the number of points of γ to which they are contiguous. If I_0, I_n, and I^* denote the interiors of $(VP_1 + VA_1)$, $(P_nT + A_kT)$, and $(TV + VA_1T)$, respectively, where TV is that arc of J not containing P_1, then*

$$I = \text{segment } VA_1T + I^* + \sum_{j=0}^{n} I_j + \sum_{j=1}^{n} \sum_{t=1}^{kj} I_{jt}.$$

If $\beta = P_1 + P_2 + P_3 + \cdots$ and $\gamma = A_1 + A_2 + A_3 + \cdots$ are infinite sets each having T as its only limit point, then

$$I = \text{segment } VA_1T + I^* + \sum_{j=0}^{\infty} I_j + \sum_{j=1}^{\infty} \sum_{t=1}^{kj} I_{jt}.$$

A similar formula holds for the case in which each of the points V and T is approached sequentially by a sequence from β and by a sequence from γ.

THEOREM 10. *Let the following changes be made in the hypotheses of Theorem 7: P_1 is contiguous to both A_1 and A_k but to no other point of γ. Either each of the points A_1 and A_k is contiguous to some point of $\beta - P_1$ or else $k > 2$. Then either it is true that if P_j and $P_{j'}$ are contiguous to A_i and $A_{i'}$, respectively, and if $j < j'$, then $i \leqq i'$, or it is true that if P_j and $P_{j'}$ are contiguous to A_i and $A_{i'}$, respectively, and if $j < j'$, then $i \geqq i'$. If the former condition holds, if I^* denotes*

the interior of $A_1A_k+P_1$, and if I_n denotes the point set obtained by substituting n for j in the definition of I_j given in the statement of Theorem 7, then

$$I = A_1A_k + I^* + \sum_{j=1}^{n} I_j + \sum_{j=2}^{n}\sum_{t=1}^{kj} I_{jt}.$$

The theorem is illustrated by Fig. 2.

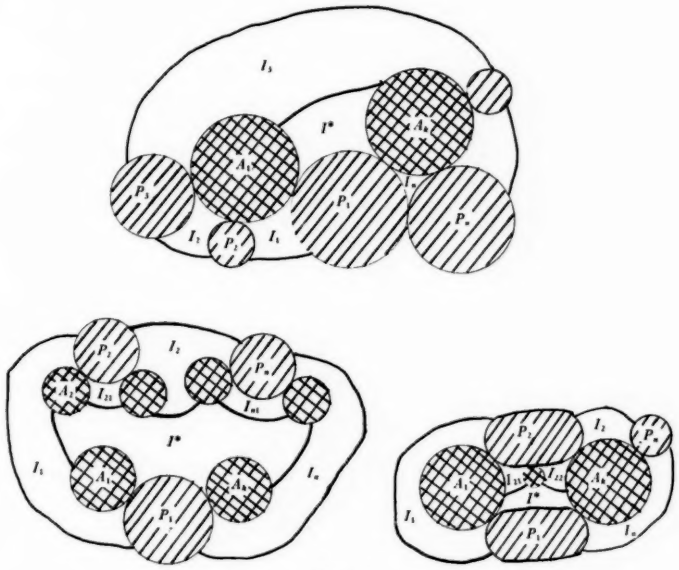


FIG. 2

If in Theorem 10, $\beta = P_1+P_2$, where P_1 and P_2 are non-contiguous, then J is the sum of two arcs, P_1XP_2 and P_1YP_2, and is such that the corresponding segments are mutually separated. Suppose $k \geqq 3$. (1) If at least one of the points A_1 and A_k, say A_1, is contiguous to P_2, then by Theorem 3, $I-A_1 = D_1+D_2$, where D_1 and D_2 are the interiors of the simple closed curves $P_1XP_2+A_1$ and $P_1YP_2+A_1$, respectively. Let I_1 denote that domain above which does not contain A_k. Suppose $I_1 = D_1$. Let I_2 denote the interior of $P_1YP_2+A_k$ or of $P_1YP_2+A_{k-1}A_k$, according as A_k is or is not contiguous to P_2. (2) If neither A_1 nor A_k is contiguous to P_2, then $I-A_1A_2 = D^*+D^{**}$, where D^* and D^{**} are the interiors of $P_1XP_2+A_1A_2$ and $P_1YP_2+A_1A_2$, respectively. Let I_1 denote the domain above which does not contain A_k. Suppose $I_1 = D^*$. Let I_2 denote the interior of $P_1YP_2+A_{k-1}A_k$. Then in either of

these two cases we have, in accordance with the above notation, the following theorem:

THEOREM 11. $I = A_1A_k + I^* + I_1 + I_2 + \sum_{t=1}^{k_2} I_{2t}$.

This theorem is illustrated by Fig. 3.

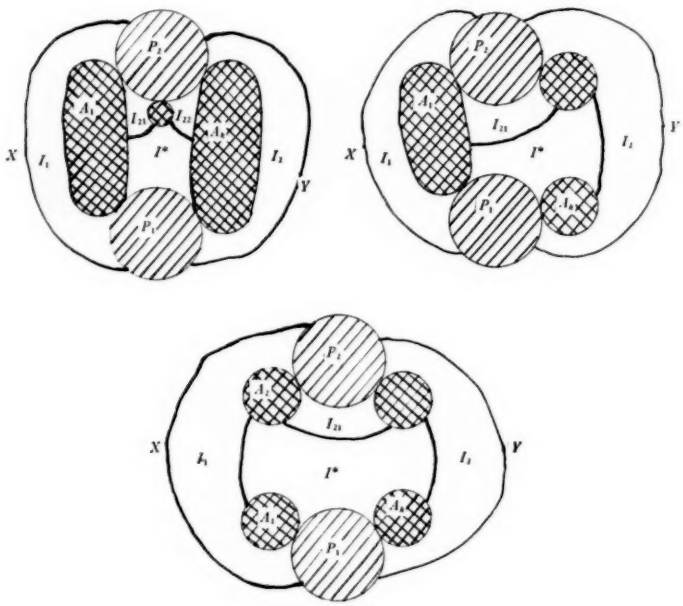


FIG. 3

THEOREM 12. *Let J and C denote two point sets, each of which is either a simple closed curve or a triune. Let I and D denote the interiors of J and C, respectively, and suppose P is a point of $I \cdot D$. Then there exists a simple closed curve or triune Q such that* (1) *Q is a subset of $J+C$, and* (2) *the interior of Q contains P and is a subset of $I \cdot D$.*

Indication of proof. If at least one of the two sets J and C is a triune, one of the sets is a subset of the other plus its interior and hence will have the properties required of Q.

A similar situation exists if both J and C are simple closed curves and one of them is a subset of the other plus its interior.

If neither of the simple closed curves J and C is a subset of the other plus its interior, let A be a point of $J \cdot D$, and let PA be an arc from P to A lying

wholly in D. Let O denote the first point, in the order P to A, that PA has in common with J. I wish to show the existence of a simple closed curve or triune Q' such that (1) Q' is a subset of $C+J\cdot D$, (2) $Q'\cdot C$ is connected, (3) $Q'\cdot J\cdot D$ is connected and contains O, and (4) the interior of Q' contains $(PO-O)$. Obviously, the existence of Q' will be established if it can be shown that there exists a point set K (which shall contain $Q'\cdot J\cdot D$) which is a connected subset of J containing O and which satisfies with respect to C those conditions satisfied by any one of the point sets P, A_1A_k, segment AXB, (A_1T-T), segment VA_1T, A_1A_k, or A_1A_k with respect to the corresponding given simple closed curve of Theorems 6, 7, 3, 8, 9, 10, and 11, respectively. There are several cases.

Case 1. Suppose O is not contiguous to any point of C. Let B denote a point of $J\cdot E$. The set J is the sum of two arcs from O to B, say OXB and OYB. Let M denote the closed point set consisting of C together with all points Z of D such that Z is contiguous to at least one point of C. Let A_1 and A_2 denote the first points that OXB and OYB have in common with M, in the order O to B. If both A_1 and A_2 are points of C, the hypothesis of Theorem 3 is satisfied, and the segment A_1OA_2 is the desired point set K.

Case 2. Suppose O is not contiguous to any point of C and that one of the points A_1 and A_2 described in Case 1, say A_1, is a point of D, while the other, A_2, is a point of C. If A_1 is contiguous to two or more points of C, then the hypotheses of Theorem 8 are satisfied, and $(A_1OA_2-A_2)$ is the desired point set K. If A_1 is contiguous to only one point F of C, and F is not contiguous to A_2, then the hypotheses of Theorem 3 are satisfied, and $(A_1OA_2-A_2)$ is the desired point set K. If F is contiguous to A_2, let A_1' denote the first point of OXB, in the order O to B, which is either a point of D that is contiguous to a point of $C-F$ or else a point of C. Now if A_1' is a point of D, the hypothesis of Theorem 8 is satisfied, and $(A_1'OA_2-A_2)$ is the desired point set K. If A_1' is a point of C, the hypothesis of Theorem 9 is satisfied, and the segment A_1OA_2 is the desired point set K.

The remaining cases may be treated with similar methods.

If Q' is a triune, Q' has the properties required of Q, since the interior of Q' cannot contain any points of J. (No simple closed curve may contain a point of each complementary domain of a triune.)

If Q' is a simple closed curve whose interior contains no point of J, then $Q'=Q$. If the interior of Q' contains any points of J, let M denote the set of all such points of J. Now, $M+Q'$ contains a simple closed curve Q having the required properties. The proof of this last statement is little different from the proof for spaces in which there do not exist contiguous points.

As an immediate result of Theorem 12 we have the following theorem:

THEOREM 13. *Let $J_1, J_2, \cdots, J_n$ denote n point sets each of which is either a simple closed curve or a triune; and suppose that $I_1, I_2, \cdots, I_n$, the interiors of $J_1, J_2, \cdots, J_n$, respectively, have a point O in common. Then there exists a simple closed curve or triune J which is a subset of $J_1+J_2+\cdots+J_n$ and is such that I, the interior of J, is a subset of $I_1 \cdot I_2 \cdot \cdots \cdot I_n$ and contains O.*

THEOREM 14. *Suppose each of the sets J_1 and J_2 is either a simple closed curve or a triune. Let I_1 and I_2 denote the interiors of J_1 and J_2, respectively, and suppose I_1+I_2 is connected. Then there exists a simple closed curve or triune J which is a subset of J_1+J_2 and whose interior is a subset of $I_1 \cdot I_2$.*

Indication of proof. Assume the theorem false. Then by Theorem 12 $I_1 \cdot I_2 = 0$. Hence there exist points P_1 and P_2 of I_1 and I_2, respectively, such that P_1 is contiguous to P_2. Thus P_1 is a point of J_2 and P_2 is a point of J_1. There exists a point F of $J_1 \cdot J_2$ and an arc P_1F of J_2 such that $P_1F - F$ is a subset of I_1. Let P_2F denote an arc of J_1. The point set P_1F+P_2F contains a simple closed curve or triune J' containing P_1 and P_2, whose interior I' is a subset of I_1. The set $J'+I' \cdot J_2$ contains a simple closed curve or triune J satisfying the required conditions. Thus the assumption is false, and the theorem is established.

With the help of Theorems 12 and 14, the following theorem may be established:

THEOREM 15. *Let J_1 and J_2 denote two point sets, each of which is either a simple closed curve or a triune. Let I_1 and I_2 denote the interiors of J_1 and J_2, respectively. Suppose I_1+I_2 is connected. Then there exists a simple closed curve or triune J which is a subset of J_1+J_2, and whose interior contains I_1+I_2.*

THEOREM 16. *Let $J_1, J_2, \cdots, J_n$ denote n point sets each of which is either a simple closed curve or a triune. Let $I_1, I_2, \cdots, I_n$ denote the interiors of $J_1, J_2, \cdots, J_n$, respectively. Suppose $I_1+I_2+\cdots+I_n$ is connected. Then there exists a point set J which is either a simple closed curve or a triune, which is a subset of $J_1+J_2+\cdots+J_n$, and which is such that I, the interior of J, contains $I_1+I_2+\cdots+I_n$.*

THEOREM 17. *If J is a simple closed curve or triune, then I, the interior of J, contains at least one point which is not contiguous to any point of J.*

Proof. Suppose each point of I is contiguous to some point of J. Let P be any point of I. I wish to show the existence of two distinct points of I, each contiguous to P. Since I is connected and contains infinitely many points, P is either contiguous to a point of I or is a limit point of I. By Axiom C, J contains all limit points of I; hence P must be contiguous to a point Q of I. In case there exists a point P_1 of J which is contiguous to both

P and Q, then I_1, the interior of the triune PQP_1, is a connected subset of I. Since J contains all the limit points of I_1, it follows that P is contiguous to a point T of I_1. Thus for this case there exist two distinct points (Q and T) of I, each contiguous to P. In case no point of J is contiguous to both P and Q, let P_1 and Q_1 denote points of J which are contiguous to P and Q, respectively, and let P_1Q_1 denote an arc of J. The point set P_1Q_1+P+Q contains a simple closed curve J_2, which contains P and Q. Let I_2 denote the interior of J_2. The argument used above may be used here to show that P is contiguous to a point T of I_2.

Let QT denote an arc lying in I. The point set $QT+P$ contains a simple closed curve or triune J' whose interior I' contains infinitely many points, no one contiguous to any point of J. Thus we reach a contradiction, and the theorem is established.

Example 1. In the euclidean plane let C_1, C_2, and C_3 denote three circles, each tangent externally to each of the others. Denote their centers by P_1, P_2, and P_3, respectively, and their radical center by O. Let $A_1=C_1\cdot C_2$, $A_2=C_2\cdot C_3$, and $A_3=C_3\cdot C_1$. Let B_1, B_2, and B_3 denote points on the rays OA_1, OA_2, and OA_3, respectively, such that $d(O, B_1)=d(O, B_2)=d(O, B_3)$, and $d(O, B_1)>d(O, A_1)$. Let B_3B_1, B_1B_2, and B_2B_3 be circular arcs having P_1, P_2, and P_3, respectively, as centers and lying on the non-O sides of the lines B_3B_1, B_1B_2, and B_2B_3, respectively. Let $\beta_{11}=B_3B_1+OB_3+OB_1$, $\beta_{21}=B_1B_2+OB_1+OB_2$, $\beta_{31}=B_2B_3+OB_2+OB_3$. For each point P of β_{i1} and each j, ($i=1, 2, 3$; $j=1, 2, 3, \cdots$), let Q_{Pij} denote the point of the interval P_iP such that $d(P_i, Q_{Pij})=d(P_i, A_i)+[d(P_i, P)-d(P_i, A_i)]/j$, and let β_{ij} denote the set of all points Q_{Pij}. Let T_i, ($i=1, 2, 3$), denote C_i plus its interior. Let S denote the following collection: (1) T_i, ($i=1, 2, 3$), is an element of S, and (2) each point of the plane not in $T_1+T_2+T_3$ is an element of S. For each positive integer n, let G_n denote the following subsets of S: (1) the interior of each circle of radius $1/n$ or less which neither encloses nor contains a point of $T_1+T_2+T_3$ is an element of G_n; (2) for each pair (i, j), ($i=1, 2, 3$; $j=n, n+1, \cdots$), the set consisting of T_i and all elements of S enclosed by β_{ij} is an element of G_n. If each element of S is called a "point," each element of G_1 is called a "region," and each of the "points" T_1, T_2, and T_3 is "contiguous to" each of the others, then each of the axioms of this paper is non-vacuously satisfied.

Example 2. In a euclidean space let s_1, s_2, s_3, s_4, and s_5 denote the spheres whose equations are $(x-1)^2+y^2+z^2=3/2^2$, $(x+1/2)^2+(y-(3)^{1/2}/2)^2+z^2=3/2^2$, and $(x+1/2)^2+(y+(3)^{1/2}/2)^2+z^2=3/2^2$, $x^2+y^2+(z-(3)^{1/2}/12)^2=3/12^2$, and $x^2+y^2+(z+(3)^{1/2}/12)^2=3/12^2$, respectively. Each of these spheres is tangent externally to each of the others. For each i, ($i=1, \cdots, 5$), let T_i denote s_i

plus its interior. Let S denote the collection defined as follows: (1) For each i, $(i=1, \cdots, 5)$, T_i is an element of S. (2) Let K denote the (x, y)-plane. Each point of the unbounded component of $K-K(T_1+T_2+T_3)$ is an element of S. (3) For each set of three distinct positive integers i, j, and k, such that each integer is less than 6 and at least one integer is either 4 or 5, let M_{ijk} denote the plane which contains the centers of the spheres s_i, s_j, and s_k. Each point of the bounded component of $M_{ijk}-M_{ijk}(T_i+T_j+T_k)$ is an element of S. For each positive integer n, let G_n denote the subsets of S defined as follows: (1) The interior of each circle of radius $1/n$ or less which is a subset of $S \cdot M_{ijk}$ or of $S \cdot K$ is an element of G_n. (2) In each of the sets $S \cdot M_{ijk}$, construct three sequences of segments like the three sequences of segments constructed within the triune of Example 1, and in $S \cdot K$ construct three sequences of segments like those constructed in the exterior of the triune of Example 1. For each pair (i, j), $(i=1, \cdots, 5; j=n, n+1, \cdots)$, the subset of S consisting of T_i together with all elements of $S-(T_1, T_2, \cdots, T_5)$ each of which is enclosed by T_i plus the jth segment of one of the sequences constructed is an element of G_n. If each element of S is called a "point," each element of G_1 is called a "region," and each of the "points" $T_1, \cdots, T_5$ is "contiguous to" each of the others, then all the axioms of this paper are non-vacuously satisfied.

Example 2 shows that Theorem 6 fails to hold if J denotes a triune instead of a simple closed curve.

THEOREM 18. *If J is a simple closed curve, if I is the interior of J, and if H and K are mutually exclusive compact continua lying in $J+I$, then no two points of H separate two points of K on J.*

Proof. Suppose on the contrary that the points A and B of H separate the points C and D of K on J. It follows from Theorem 24 of S.C.P. that K contains an irreducible continuum T from $K \cdot ACB$ to $K \cdot ADB$. By Theorem 28 of S.C.P., $T-T \cdot J$ is a connected set having boundary points C_1 and D_1 in segments ACB and ADB, respectively. It may be readily shown that the component of $S-(J+H)$ which contains $T-T \cdot J$ contains a segment C_2YD_2, where C_2 and D_2 are points of segments ACB and ADB, respectively, and neither point belongs to H. Thus the arc C_2YD_2 contains no point of H and lies, except for its end points, in I.

The above argument may be repeated to show that there exists an arc A_2XB_2 which contains no point of C_2YD_2, where the segment A_2XB_2 lies in I, and A_2 and B_2 separate C_2 and D_2 on J. But this contradicts Theorem 4.

THEOREM 19. *The interior of a simple closed curve or triune is not a subset of any simple closed curve or triune.*

Proof. Let I denote the interior of a simple closed curve or triune J. By Theorem 5, I is not a subset of a triune. Suppose I is a subset of a simple closed curve C. By Theorem 17, there exists a point P of I which is not contiguous to any point of J. Now P is either a limit point of $S-C$ or is contiguous to a point of $S-C$. If P is a limit point of $S-C$, let R be a connected domain containing P, but containing no point of J, and let Q be a point of $R\cdot(S-C)$. If P is not a limit point of $S-C$, let Q be a point of $S-C$ which is contiguous to P. In either case, Q belongs to $I\cdot(S-C)$.

THEOREM 20. *If neither of the contiguous points X and Y separates the point A from the point B, then their sum does not separate A from B.*

Proof. Suppose $S-(X+Y)=S_A+S_B$, where S_A and S_B are mutually separated sets containing A and B, respectively. There exists an arc α from A to B which does not contain Y. The arc α contains X; hence α contains an arc AX. Similarly, there exists an arc AY. The point set $AX+AY-(X+Y)$ is connected and hence lies in S_A. The set $AX+AY$ contains a simple closed curve or triune J which contains X and Y and is a subset of S_A+X+Y. Let Q be any point of $J-(X+Y)$, and let I and E denote the two complementary domains of J. Since Q is a point of S_A and is a boundary point of each of the connected sets I and E, it follows that I and E are subsets of S_A. Hence S is a subset of S_A+X+Y. Thus we reach a contradiction, and the theorem is established.

AXIOM 4. *If P and Q are two distinct non-contiguous points, there exists a simple curve or triune which separates P from Q.*

With the help of Axiom 4, the Borel-Lebesgue theorem (Theorem 5 of S.C.P.), and Theorem 16, the next theorem may be established.

THEOREM 21. *If H and K are two mutually separated compact continua, there exists a simple closed curve or triune which separates H from K.*

THEOREM 22. *If J is a simple closed curve or triune containing the contiguous points A and B, I is the interior of J, and P is a point of $J+I$ which is not contiguous to A, then there exists a simple closed curve or triune J^* satisfying the following conditions*: (1) *It is a subset of $J+I$.* (2) *Its intersection with J is an arc containing A and B.* (3) *Its exterior contains P.*

Proof. Suppose the theorem is false. If may be readily shown that there exists a simple closed curve or triune which satisfies conditions (1) and (2). Furthermore it may be readily shown that if P is any point of J not contiguous to A, or if P is a point of I not contiguous to A such that there is a simple closed curve or triune J' containing P and satisfying conditions (1) and (2),

then there exists a simple closed curve or triune satisfying conditions (1), (2), and (3). Hence our supposition implies that P is a point of I, and every simple closed curve or triune, satisfying conditions (1) and (2), encloses P. Let PX denote an arc from P to some point X of $J-(A+B)$ such that $PX-X$ is a subset of I. Let S_1 denote the set consisting of P and all points Y of $PX-X$ such that every simple closed curve or triune satisfying conditions (1) and (2) encloses the interval PY of PX, and let S_2 denote $PX-S_1$. By the Dedekind-cut proposition (P.S.T., chap. 1, Theorem 64) there exists a point Q which is either the last point of S_1 or the first point of S_2, in the order P to X.

Suppose Q is the first point of S_2. There exists a simple closed curve or triune J' satisfying the following conditions: If $Q=X$, then $J'=J$; if $Q\neq X$, then J' contains Q, satisfies conditions (1) and (2), and encloses $PQ-Q(=S_1)$. There exists a third simple closed curve or triune J_1, which is a subset of J' plus its interior and is such that $J'\cdot J_1$ is an arc containing A and B but not containing Q. Thus Q is exterior to J_1; hence $PQ-Q$ either intersects J_1 or is exterior to J_1. Both these possibilities are ruled out since J_1 satisfies conditions (1) and (2) and hence encloses $PQ-Q$. Thus Q is not the first point of S_2.

Suppose Q is the last point of S_1. By means of an argument analogous to that used in the last paragraph, it may be shown that Q is not contiguous to a point of S_2. Thus Q is a limit point of S_2. Now Q cannot be contiguous to both A and B; for if such were the case, the triune ABQ would satisfy conditions (1) and (2) and hence would enclose $PQ=S_1$ and, in particular, the point Q. Let C denote a point of the pair (A, B) which is not contiguous to Q, and let $C\omega$ be an arc from C to ω lying in J plus its exterior and containing no point of J which is contiguous to Q. By Theorem 21, there exists a simple closed curve or triune J' which separates Q from $C\omega$. Since Q is a limit point of S_2, it follows that I', the interior of J', contains a segment QW of S_2. There exist a point F of the segment QW, and a simple closed curve or triune J_1 which satisfies conditions (1) and (2) and is such that F is in E_1, the exterior of J_1. Let I_1 denote the interior of J_1. By Theorem 12, there exists a simple closed curve or triune J_2, which is a subset of J_1+J' and whose interior I_2 contains Q and is a subset of $I_1\cdot I'$. There exists a point T of $J_2\cdot I'\cdot J_1$. It may be readily shown that the point set $J_1-T+(J_2-T)$ contains a simple closed curve or triune J^* which satisfies conditions (1) and (2). Furthermore, since I_2+T+E_1 is a connected point set containing Q and ω but no point of J^*, Q is exterior to J^*. Hence Q is not a point of S_1.

Thus we have reached a contradiction and the theorem is established.

THEOREM 23. *The interior of every triune is non-compact and so is the interior of every simple closed curve which contains two contiguous points.*

Proof. Let J denote a simple closed curve or triune containing the contiguous points A and B, and let I denote the interior of J. Suppose that I is compact. It follows that $J+I$ is compact, completely separable, and metric. Let $D_1, D_2, D_3, \cdots$ be a sequence of domains which properly covers $J+I$ and with respect to which $J+I$ is completely separable. Let n_1 denote the smallest integer (which exists in view of Theorems 17 and 22) such that there exists a simple closed curve or triune J_1 which is a subset of $J+I$, and such that $J \cdot J_1$ is an arc containing A and B, and D_{n_1} is exterior to J_1. Let $n_1, n_2, n_3, \cdots$ be an increasing sequence of positive integers satisfying the following conditions: For each integer $k>1$, n_k is the smallest integer greater than n_{k-1} such that there exists a simple closed curve or triune J_k which is a subset of J_{k-1} plus its interior, and such that $J_{k-1} \cdot J_k$ is an arc containing A and B, and D_{nk} is exterior to J_k. For each integer r, let α_r denote the point $J_r-(A+B)$ in case J_r is a triune. Otherwise, let α_r denote an arc P_rQ_r of $J_r-(A+B)$, where Q_r is either contiguous to B or else $d(Q_r, B)<1/r$, and where P_r is either contiguous to A or else $d(P_r, A)<1/r$. For each integer n let E_n denote the exterior of J_n. For each integer r and each point P of α_r there exists an integer n_P such that E_{n_P} contains P. By the Borel-Lebesgue theorem, there exists a finite collection of these domains covering α_r. The domain of this finite collection with greatest subscript E_s, contains all other domains of the finite collection; therefore E_s contains α_r. Hence $\alpha_r \cdot J_s=0$, and consequently $\alpha_r \cdot \alpha_s=0$. Thus there exists an increasing sequence of integers $r_1, r_2, r_3, \cdots$ such that for each integer $k>0$, $\alpha_{r_k} \cdot \alpha_{r_{k+1}}=0$; hence $\alpha_{r_1}, \alpha_{r_2}, \cdots$ is a sequence of mutually exclusive continua. There exists a subsequence of this sequence which converges to a sequential limiting set L containing A and B. Hence by Theorem 33 of S.C.P., L is a perfect continuum and is therefore uncountable. For each integer n, L is a subset of J_n plus its interior. Hence if L contains a point X other than A or B, then X is contiguous to A or B. But the set of all such points X is at most countable by Theorem 14 of S.C.P. Hence L is countably infinite or finite. We have thus reached a contradiction, and the theorem is established.

THEOREM 24.* *If H and K are two mutually separated, closed and compact point sets containing the points A and B, respectively, then there exists a simple closed curve or triune J which separates A from B such that $J \cdot (H+K)=0$.*

Proof. Let h_A and k_B denote the components of H and K containing A and B, respectively. From Theorem 21 it follows that for each component h of H there exists a simple closed curve or triune J_{hB} which separates h from

* Cf. L. Zoretti, *Sur les fonctions analytiques uniformes*, Journal de Mathématiques Pures et Appliquées, vol. 1 (1905), pp. 9–11.

k_B. Let D_{kB} denote that complementary domain of J_{kB} which contains h. The collection of all such domains D_{kB} covers H; hence there exists a finite collection T of such domains covering H. Let T^* denote the point set which is the sum of all the elements of T. By Theorem 16 there exists a simple closed curve or triune J_{AB} whose interior with respect to B contains that component of T^* which contains A. Furthermore, $(H+k_B)\cdot J_{AB}=0$. For each component k of K let J_{Ak} denote a simple closed curve or triune (which exists in view of the preceding argument) which separates A from k and is such that $(H+K)\cdot J_{Ak}=0$. Let D_{Ak} denote the interior of J_{Ak} with respect to A. The collection of all such domains D_{Ak} covers K; hence there exists a finite collection G of such domains covering K. Let G^* denote the sum of the elements of G. By Theorem 16 there exists a simple closed curve or triune J whose interior with respect to A contains that component of G^* which contains B. Furthermore, $(H+K)\cdot J=0$.

THEOREM 25.† *If H and K are two mutually separated closed and compact point sets, and if neither H nor K separates the point A from the point B, then $H+K$ does not separate A from B.*

Proof. Suppose $H+K$ separates A from B.

Case 1. Suppose H is connected. With the help of Theorem 24, it may be shown that there exists a domain D which contains K but no point of H and is the sum of a finite number of components each of which is bounded by a simple closed curve or triune containing no point of $H+K$. Let AXB denote an arc from A to B containing no point of K, and AYB an arc containing no point of H. Let $J_1, J_2, J_3, \cdots, J_r$ denote the boundaries of those components of D which have points in common with AYB. There are four possibilities: (1) A and B are both in $S-D$, (2) A and B are the same component of D, (3) A and B are in different components of D, (4) A or B is in $S-D$ while the other is in D. In each case it may be readily shown that $(AYB-AYB\cdot D)+J_1+J_2+\cdots+J_r+AXB\cdot D$ contains an arc from A to B.

Case 2. Suppose H has only a finite number of components, $h_1, h_2, \cdots, h_n$. By Case 1, h_1+K does not separate A from B. The sets h_2 and h_1+K are mutually separated; hence by Case 1, $h_2+(h_1+K)$ does not separate A from B. Continuing this process we obtain the fact that $h_n+(h_{n-1}+\cdots+h_1+K)$ does not separate A from B.

Case 3. Suppose H has infinitely many components. Let AXB denote an arc containing no point of K. By Axiom C, at most a finite number of components of K contain points which are contiguous to points of AXB. Let

† Cf. P. Alexandroff, *Sur les multiplicités cantoriennes et le théorème de Phragmèn-Brouwer généralisé*, Comptes Rendus de l'Académie des Sciences, Paris, vol. 183, pp. 722–724.

$k_1, k_2, \cdots, k_n$ be the set of all such components of K. By Case 2 there exists an arc AYB which contains no point of $H+k_1+\cdots+k_n$. If $AYB\cdot K=0$, the theorem is established. Suppose $AYB\cdot K\neq 0$. For each component k of K which intersects AYB, the set $H+AXB+K$ is the sum of two mutually separated sets, one containing A and the other containing k, by Theorem 27 of S.C.P. By Theorem 24 there exists a simple closed curve or triune J_k which separates A from k and is such that $J_k\cdot(H+AXB+K)=0$. Let I_k denote the interior of J_k with respect to A. The collection of all such domains I_k covers the closed point set $AYB\cdot K$; hence there exists a finite subcollection T doing so. Let T^* denote the sum of the elements of T. By Theorem 16, for each component t of T^* there exists a simple closed curve or triune C_t which is a subset of the boundary of t and whose interior with respect to A contains t. Let $C_1, C_2, \cdots, C_s$ be the set of all simple closed curves or triunes thus obtained. The point set $(AYB-AYB\cdot T^*)+C_1+C_2+\cdots+C_s$ contains an arc from A to B which contains no point of $H+K$.

THEOREM 26. *If J is a simple closed curve, I is a complementary domain of J, H and K are two mutually separated subcontinua of J, α and β are the two components of $J-(H+K)$, and C is a simple closed curve which separates H from K, then there exists an arc AXB such that A and B are points of α and β, respectively, and segment AXB is a subset of $I\cdot C$.*

Proof. Suppose the theorem is false. Let ω be a point of $S-(J+I+C)$, which exists by Theorem 19. Let D denote the interior of C. One of the sets H and K, say H, is a subset of D. Obviously, C contains at least one point of each of the sets α, β, and I. Let P, Q, Z, and W denote points of $\alpha\cdot C$, $\beta\cdot C$, H, and K, respectively. Let α' and β' denote continua which are subsets of α and β, respectively, and which contain $\alpha\cdot C$ and $\beta\cdot C$, respectively. Now, $\alpha'+\beta'+I\cdot C$ is the sum of two mutually separated continua M_1 and M_2 containing α' and β', respectively. Let E denote the last point of the arc PZQ, in the order P to Q, which is either a point of M_1 or contiguous to a point of M_1; and let F denote the first point of PZQ which is either a point of M_2 or contiguous to a point of M_2. It may be readily seen with the help of Theorem 4 that if E and F are distinct, then E precedes F in the order P to Q. Suppose first that E and F are distinct non-contiguous points. The interval EF of PZQ contains at least one point of H. Furthermore, there exists an arc OL where L is a point of segment EF and $OL-L$ is a subset of $I-I\cdot C$. By Theorem 12 there exists a simple closed curve J' which is a subset of $J+C$ and contains EF, and whose interior I' is a subset of $I\cdot D$ and contains $OL-L$. Now J' contains no point of K, for otherwise $H+I'+K$ would be a connected set containing no point of C. Thus J' contains an arc $E'F'$, where E' and F'

are points of segments LEW and LFW, respectively, and the segment $E'F'$ is a subset of $I \cdot C$. Thus segment $E'F'$ is a subset of M_1 and also a subset of M_2. From this contradiction, it follows that E and F are identical or contiguous. Also $(E+F) \cdot (M_1+M_2)=0$. Similarly, if T denotes the last point of PWQ which is either a point of M_1 or contiguous to a point of M_1, and if V denotes the first point of PWQ which is either a point of M_2 or contiguous to a point of M_2, then T and V are identical or contiguous. Let C' denote a simple curve which separates M_1 from M_2. Let $P'Q'$ denote an arc such that P' and Q' are points of M_1 and M_2, respectively, and segment $P'Q'$ is a subset of $J+I-(J+I) \cdot (C+E+F+T+V)$. Let G denote a point of segment $P'Q'$ which belongs to C', and let NY denote an arc which is a subset of C' and contains G, where N and Y are points of J and segment NY is a subset of I. Now N and Y do not lie in different components of $J-(\alpha'+\beta')$, for otherwise, $H+K+NY$ would be a connected set containing no point of C. Suppose both N and Y are points of that component of $J-(\alpha'+\beta')$ which contains H. If either N or Y preceded E in the order PZQ, we would have two arcs (one a subset of $C \cdot M_1+E$, the other a subset of $C \cdot M_2+$interval $Q'G$ of arc $P'Q'+$interval NG of arc NY) satisfying the hypothesis of Theorem 4 but not the conclusion of this theorem. Similarly, it may be shown that neither N nor Y follows F. Hence N is identical with Y or contiguous to Y. Again a contradiction is reached, and the theorem is established.

THEOREM 27. *If A, X, B, and Y are points of the simple closed curve J in the order indicated, if I is the interior of J, and if H and K are mutually separated closed and compact subsets of $J+I$ such that $H \cdot AXB=0$ and $K \cdot AYB=0$, then there exists an arc AB lying in $J+I-(H+K)$.*

Proof. Two cases will be considered.

Case 1. Suppose there exists no component T of $H+K$ such that $T+J-(A+B)$ is connected. Let N_1 and N_2 denote continua which are subsets of segments AYB and AXB, respectively, and which contain $H \cdot AYB$ and $K \cdot AXB$, respectively; and let α and β denote the components of $J-(N_1+N_2)$ which contain A and B, respectively. By Theorem 27 of S.C.P., N_1+N_2+H+K is the sum of two mutually separated closed sets M_1 and M_2 containing N_1 and N_2, respectively. Hence by Theorem 24 there exists a simple closed curve C which separates N_1 from N_2 and contains no point of M_1+M_2. Thus C contains no point of $H+K$. By Theorem 26 there exists an arc $A'X'B'$, where A' and B' are points of α and β, respectively; and segment $A'X'B'$ lies in $I \cdot C$. Now $A'X'B'+\alpha+\beta$ contains an arc AB lying in $J+I-(H+K)$.

Case 2. Suppose that for at least one component T of $H+K$, $T+J$

$-(A+B)$ is connected. With the help of Axiom C, it may be shown that there is only a finite number of such components. Either H or K, say H, contains at least one such component. There exists a point Y_1 which is the first point of segment AYB, in order A to B, which is either a point of, or contiguous to a point of, such a component of H.

First suppose there exists at least one point Q of AY_1-A having the property that there is a component T_Q of K such that $Q+T_Q+$segment AXB is connected. Let $Q_1, Q_2, \cdots, Q_n$ denote all such points in the order A to Y_1. Obviously Q_i, $(i=1, 2, \cdots, n)$, is not a point of $H+K$. If the segment AQ_1 of AY_1 contains no point of H, let α_1 denote the interval AQ_1. If segment AQ_1 contains a point of H, it may be shown with the help of Theorem 18 and Case 1 of this theorem that there exists an arc α_1, from A to Q_1, lying in $J+I-(H+K)$. Similarly, if $n>1$, there exists an arc α_i, $(i=2, 3, \cdots, n)$, from Q_{i-1} to Q_i, lying in $J+I-(H+K)$. The set $\alpha_1+\alpha_2+\cdots+\alpha_n$ contains an arc α from A to Q_n. Now let X_2 denote the last point of segment AXB, in order A to B, for which there exists a component T_2 of H, such that $Y_1+T_2+X_2$ is connected. If $Q_n=Y_1$, and if there exists a point P of AXB between X_2 and B for which there is a component T' of K such that $Y_1+T'+P$ is connected, then, denoting Q_n by Q', we have an arc $AQ'=\alpha$ having the following properties: (1) The arc AQ' lies in $J+I-(H+K)$, (2) Q' is a point of segment AYB and does not precede Y_1, (3) there is at least one component T of H for which $[T+$segment $AXB+$interval AQ' (of AYB) $-A]$ is connected, but $T+Q'B-Q'$ is not connected, and (4) there exists no component W of $H+K$ such that $W+AYB-Q'$ is connected. If $Q_n=Y_1$, and if there exists no such point P, let X_1 denote the first point of AXB, such that $Y_1+T_2+X_1$ is connected; or if $Q_k\neq Y_1$, let X_1 denote the first point of AXB for which there exists a component T of H such that Y_1+T+X_1 is connected. In either case, X_1 is a point of segment AXB, and with the help of Theorem 18, Case 1, it may be shown that there exists an arc β from Q_n to X_1 lying in $J+I-(H+K)$. If $X_1\neq X_2$, then by means of an argument like that used in obtaining α, it may be shown that there exists an arc γ, from X_1 to X_2, lying in $J+I-(H+K)$. If $X_1=X_2$, let γ denote X_1. If there exists no point X' of AXB between X_2 and B for which there is a component T of K such that segment $AYB+T+X'$ is connected, then the argument for obtaining α may be used here to obtain an arc δ, from X_2 to B, lying in $J+I-(H+K)$. In this case, $\alpha+\beta+\gamma+\delta$ contains an arc from A to B, and the theorem is established. If there exists a point X', as described, let Q' denote the first point of segment AYB for which exists a component T of K, such that $Q'+T+$segment X_2B is connected. It follows from Theorem 18 and Case 1, that there exists an arc η, from X_2 to Q', lying in $J+I-(H+K)$. In this case, $\alpha+\beta+\gamma+\eta$ con-

tains an arc AQ' having the four properties listed above.

Now suppose there is no point Q as described above. Let X_2 have the same meaning as above, and let γ denote an arc from A to X_2 lying in $J+I-(H+K)$. Again, if there exists no point X', as described above, let δ denote an arc from X_2 to B lying in $J+I-(H+K)$. In this case $\gamma+\delta$ contains an arc from A to B, and the theorem is established. If there exists a point X', as described above, let Q' and η be defined as before. Now $\gamma+\eta$ contains an arc AQ' with properties listed above.

Thus, in any case either there exists an arc AB lying in $J+I-(H+K)$, or else there exists an arc AQ' having the properties listed above. In the latter case, if there exists no point Y' of AYB, between Q' and B, for which there exists a component T of H such that $Y'+T+$segment AXB is connected, then the argument for obtaining α may be used to obtain an arc $Q'B$ lying in $J+I-(H+K)$. In such a case, $AQ'+Q'B$ contains an arc from A to B, and the theorem is established. If there exists a point Y', as described above, let Y_2 denote the first such point in the order Q' to B. Now the argument for obtaining AQ' may be repeated to obtain either an arc $Q'B$ lying in $J+I-(H+K)$ (in which case $AQ'+Q'B$ contains an arc from A to B, and the theorem is established), or an arc $Q'Q''$ of such nature that $AQ'+Q'Q''$ contains an arc AQ'' having the following properties: (1) AQ'' lies in $J+I-(H+K)$, (2) Q'' is a point of segment AYB and does not precede Y_2, (3) there are at least two components of H such that if T denotes either of them, then $[T+$segment $AXB+$interval AQ'' of $AYB-A]$ is connected, but $[T+Q''B-Q'']$ is not connected, and (4) there exists no component W of $H+K$, such that $W+AYB-Q''$ is connected. Since there is only a finite number of components of H such that if T denotes any one of them, $T+J-(A+B)$ is connected, the above process may be repeated until we have obtained either an arc AB lying in $J+I-(H+K)$, or an arc $AQ^{(k)}$ lying in $J+I-(H+K)$, where $Q^{(k)}$ is a point of segment AYB, and there is no point Y' between $Q^{(k)}$ and B for which there is a component T of H such that $Y'+T+$segment AXB is connected. In the latter case, the argument for obtaining α may be used to obtain an arc $Q^{(k)}B$ lying in $J+I-(H+K)$. The set $AQ^{(k)}+Q^{(k)}B$ contains an arc AB lying in $J+I-(H+K)$; and the theorem is established.

THEOREM 28.* *If the common part of the closed and compact point sets H and K is a continuum, and if neither H nor K separates the point A from the point B, and if furthermore, $H-H\cdot K$ and $K-H\cdot K$ are mutually separated sets, then $H+K$ does not separate A from B.*

* Cf. S. Janiszewski, *Sur les coupures du plan faites par des continus*, Prace Matematyczno-Fizyczne, vol. 26, 1913. Also Anna M. Mullikin, *Certain theorems relating to plane connected point sets*, these Transactions, vol. 24 (1922), pp. 144–162.

Proof. Suppose on the contrary that $H+K$ separates A from B. Let $M=H+K$ and $T=H\cdot K$. Let S_1 and S_2 denote the components of $S-M$ which contain A and B, respectively. There exist arcs AXB and AYB such that $AXB\cdot H=0$ and $AYB\cdot K=0$. Let X_1 and Y_1 denote the first points of AXB and AYB, respectively, which belong to the boundary of S_2. Thus X_1 and Y_1 belong to $K-T$ and $H-T$, respectively, and hence are not contiguous to each other. Let AX_1 and AY_1 denote intervals of AXB and AYB, respectively. The set AX_1+AY_1 contains an arc $X_1A_1Y_1$ such that $X_1A_1\cdot H=0$ and $A_1Y_1\cdot K=0$. If X_1 is contiguous to any point of S_2, let X_2 denote such a point, and let X_1X_2 denote the arc consisting of these two points. Otherwise, let R denote a connected domain containing X_1 but containing no point or boundary point (other than X_1) of $H+A_1Y_1$, let X_2 denote a point of $R\cdot S_2$, and let X_1X_2 denote an arc lying in R. If Y_1 is contiguous to any point of S_2, let Y_2 denote such a point, and let Y_1Y_2 denote the arc consisting of these two points. Otherwise, let W denote a connected domain containing Y_1 but containing no point or boundary point (other than Y_1) of $K+A_1X_1+X_1X_2$, let Y_2 denote a point of $W\cdot S_2$, and let Y_1Y_2 denote an arc lying in W. Let X_2Y_2 denote either the point X_2 or an arc lying in S_2 according as X_2 is or is not Y_2. It may be readily shown that the set $X_1A_1Y_1+X_1X_2+Y_1Y_2+X_2Y_2$ contains a simple closed curve J such that (1) J contains A_1 and a point B_1 of X_2Y_2, and (2) of the two segments of J from A_1 to B_1, one contains no point of H and the other contains no point of K. Let I denote that complementary domain of J which does not contain T, and let $H_1=H\cdot(J+I)$ and $K_1=K\cdot(J+I)$. It follows from Theorem 27 that there exists an arc A_1B_1 lying in $J+I-(H_1+K_1)$. Thus A_1B_1 contains no point of M. But this is impossible since A_1 and B_1 lie in different complementary domains of M. The theorem is thus established.

THEOREM 29. *If no point of the arc XY separates the point A from the point B, then XY does not separate A from B.*

Proof. Suppose $S-XY=S_A+S_B$, where S_A and S_B are mutually separated sets containing A and B, respectively. Let S_1 denote the set consisting of X together with all points Z, if there are any, such that the interval XZ does not separate A from B, and let $S_2=XY-S_1$. Now S_2 contains Y, and clearly every point of S_1 precedes every point of S_2. Hence there exists either a last point of S_1 or a first point of S_2.

Suppose there exist a point O which is the last point of S_1 and Q which is the first point of S_2. By Theorem 20 the interval OQ does not separate A from B. Hence by Theorem 28, S_1+OQ does not separate A from B. Thus XQ does not separate A from B, contrary to the fact that Q is a point of S_2.

Suppose that there exists a point O which is the last point of S_1. Let $O_1, O_2, O_3, \cdots$ be a sequence of points of S_2 converging to O such that each precedes the next in the order Y to X. By Theorem 28, the interval OO_n separates A from B. Thus, by Theorem 5, Chapter 2, of P.S.T., the point O separates A from B.

Suppose there exists a point O which is the first point of S_2. Let $O_1, O_2, O_3, \cdots$ be a sequence of points of S_1 converging to O such that each precedes the next in the order X to Y. By Theorem 28, the interval O_nO separates A from B. Hence by Theorem 5, Chapter 2, of P.S.T., the point O separates A from B. Again a contradiction is reached and the theorem is established.

The following proposition is false:

PROPOSITION. *If A, B, and C are three distinct points, and if A is not contiguous to either B or C, then there exists a simple closed curve or triune which separates A from $B+C$.*

In Example 1, let A be any point of the triune $T_1T_2T_3$, and let B and C be points of different complementary domains of this triune. There does not exist a simple closed curve or triune which separates A from $B+C$.

AGRICULTURAL AND MECHANICAL COLLEGE OF TEXAS,
COLLEGE STATION, TEXAS

ON INTEGRATION IN VECTOR SPACES*

BY

B. J. PETTIS

Introduction. Several authors ([3]–[10], inclusive; [15])† have already given generalized Lebesgue integrals for functions $x(s)$ whose values lie in a Banach space (B-space) $\mathfrak{X}$.‡ In the following pages another definition,§ based on the linear functionals over $\mathfrak{X}$ and on the ordinary Lebesgue integral,‖ will be given and its properties and relationships to other integrals discussed.

We consider functions $x(s)$ defined to a B-space $\mathfrak{X}=[x]$ from an abstract space $S=[s]$ possessing both a σ-field Σ of "measurable" sets having S as an element and a non-negative bounded c.a. (completely additive) "measure" function $\alpha(E)$ defined over Σ. Notational conventions are as follows: $\overline{\mathfrak{X}}$ denotes the B-space conjugate to $\mathfrak{X}$ ([1], p. 188), $f=f(x)$, $g=g(x)$, $\cdots$ denote elements of $\overline{\mathfrak{X}}$, and $F=F(f)$, $G=G(f)$, $\cdots$ elements of $\overline{\overline{\mathfrak{X}}}$, the conjugate space of $\overline{\mathfrak{X}}$; and for real-valued functions Greek letters will be used. When the abstract functions $x(s)$, $y(s)$, $\cdots$, or the real functions $\phi(s)$, $\psi(s)$, $\cdots$ are considered as elements of a functional space, they will sometimes be written $x(:)$, $y(:)$ or $\phi(:)$, $\psi(:)$. The "zero" element of $\mathfrak{X}$ will be denoted by θ.

The contents of the paper may be outlined as follows. The first section compares measurability of functions under the strong and weak topologies of $\mathfrak{X}$. The second defines the ($\mathfrak{X}$) integral and utilizes an essential lemma, first stated by Orlicz, to prove the complete additivity and absolute continuity of the integral. The linear operations from L^p to $\mathfrak{X}$ and from $\overline{\mathfrak{X}}$ to $\overline{L^p}$ defined by integrable functions are investigated in §3, and these results are expressed in terms of real-valued kernels in §7. Approximation and convergence theorems occupy §4 and lead to §5 and the relationships between the ($\mathfrak{X}$) integral and the integrals given by other definitions. A few remarks on completely continuous transformations and on differentiation account for §§6 and 8, respectively, and four examples form §9. In conclusion a few open questions are cited.

* Presented to the Society, March 27, 1937; received by the editors August 28, 1937.

† Numerals in brackets refer to the list of references at the end of the paper.

‡ See [1]. The terminology used will be that of Banach's book.

§ Suggested to the author by E. J. McShane. The definition is also due to Dunford, *Uniformity in linear spaces*, these Transactions, vol. 44 (1938), pp. 305–356. Also see [10], p. 50, footnote.

The author wishes to take this opportunity to express his gratitude to Professor McShane for for his most kind and helpful criticism during the writing of this paper.

‖ From the definition it will be evident that any integral for real-valued functions can be generalized in the same fashion.

For reference purposes the following well known theorems concerning B-spaces will be listed:*

A. *Given a linear functional (an additive, continuous, real-valued function) $g(y)$ defined in a linear subset $\mathfrak{Y}$ of $\mathfrak{X}$, there exists a linear functional $f(x)$ defined in $\mathfrak{X}$ and such that $f(y)=g(y)$, for y in $\mathfrak{Y}$, and $\|f\|=\|g\|$ where the norm of f is taken over $\mathfrak{X}$ and that of g over $\mathfrak{Y}$.*

B. *For every x_0 in $\mathfrak{X}$ there is a linear functional $f_0(x)$ defined over $\mathfrak{X}$ and such that $|f_0(x_0)|=\|x_0\|$, and $\|f_0\|=1$.*

C. *Given a sequence $\{x_n\}$ of elements of $\mathfrak{X}$, the closed linear hull (the closure of the set of all finite linear combinations) of the x_n's is a separable sub-B-space of $\mathfrak{X}$.*

1. **Measurable and weakly measurable functions.** A function $x(s)$ from S to $\mathfrak{X}$ will be called a *step-function* if and only if $x(s)$ is constant on each of a finite number of disjoint measurable sets whose sum is S. As in Bochner [4], $x(s)$ is called *measurable* provided it is a.e. (almost everywhere) the strong limit of a sequence of step-functions; a measurable function taking only a countable number of distinct values will be said to be *countably-valued*. If there exists a separable sub-B-space $\mathfrak{Y}$ and a set S' such that $\alpha(S')=\alpha(S)$ and s in S' implies $x(s)$ in $\mathfrak{Y}$, then $x(s)$ is *separably-valued*.

It is evident that a measurable function $x(s)$ is also *weakly measurable*, that is, $f(x(s))$ is measurable ([2], p. 251) for every f in $\bar{\mathfrak{X}}$. The relationship between measurability and weak measurability is given in the following theorem:

THEOREM 1.1. *A necessary and sufficient condition that $x(s)$ be measurable is that it be weakly measurable and separably-valued.*

If $x(s)$ is the limit a.e. of step-functions $x_n(s)$, then almost all of its values lie in the separable closed linear hull (see C) of the denumerable set of values of the functions $x_n(s)$; thus $x(s)$ is separably-valued.

Now suppose $x(s)$ is weakly measurable and, with no loss of generality, that all the values of $x(s)$ lie in a separable subspace $\mathfrak{Y}$.

The first conclusion is that $\|x(s)\|$ is measurable. Let $\{g_i\}$ be weakly dense ([1], p. 123) in the surface $\mathfrak{S}$ of the unit sphere of $\bar{\mathfrak{Y}}$. If $A\equiv S[\|x(s)\|\leqq a]$, and $A_g\equiv S[|g(x(s))|\leqq a]$ for each g in $\mathfrak{S}$, then $A=\prod A_g$, where the product is taken over $g\in\mathfrak{S}$. For if $\|x(s)\|\leqq a$, then for any g in $\mathfrak{S}$ we have $\|g\|=1$ and hence $|g(x(s))|\leqq\|x(s)\|\leqq a$; and for the fixed element $x(s)$ of $\mathfrak{Y}$ there is, by B, a $\bar{g}$ in $\mathfrak{S}$ such that $\|x(s)\|=|\bar{g}(x(s))|$. Thus $A=\prod A_g\subset\prod_i A_{g_i}$, where the first

* The first two are Theorems 2 and 3, [1], p. 55.

product is taken over $g \in \mathfrak{S}$. The sequence $\{g_i\}$ being weakly dense, for any g in $\mathfrak{S}$ there is a subsequence $\{g_i'\}$ such that $g_i'(x) \to g(x)$ for every x in $\mathfrak{X}$; thus if s is in $\prod A_{g_i}$, then $|g(x(s))| = \lim_i |g_i'(x(s))| \leqq a$ for each g in $\mathfrak{S}$, so that $A = \prod A_g = \prod_i A_{g_i}$, where the first product is taken over $g \in \mathfrak{S}$. From A and the hypotheses on $x(s)$, each set A_{g_i} is measurable, which implies measurability for $\|x(s)\|$.

If each $x(s)$ is covered with an open $1/n$-sphere in separable $\mathfrak{Y}$, by Lindelöf's theorem a countable set $\mathfrak{Y}_{in}$, $(i = 1, 2, \cdots, j, \cdots)$, of these spheres contains the set of functional values of $x(s)$. If x_{in} is the center of $\mathfrak{Y}_{in}$, then $x(s) - x_{in}$ is weakly measurable and has all its values in $\mathfrak{Y}$, so that $\|x(s) - x_{in}\|$ is measurable. Hence the set E_{in} composed of all s having $x(s)$ in $\mathfrak{Y}_{in}$ is a measurable set, and $\sum_i E_{in} = S$. If $x_n(s) = x_{in}$ for s in $E_{in}^* \equiv E_{in} - \sum_{j=1}^{i-1} E_{jn}$, then $\|x(s) - x_n(s)\| < 1/n$ for all s, since $\sum_i E_{in}^* = S$. Since the sets E_{jn} and E_{in}^* are measurable, $x_n(s)$ is constant on each of a countable number of disjoint measurable sets whose sum is S; $x_n(s)$ is therefore measurable, hence $x(s)$, the uniform limit of $\{x_n(s)\}$, is also measurable.

COROLLARY 1.11. *If $\mathfrak{X}$ is separable, measurability is equivalent to weak measurability.*

COROLLARY 1.12. *A function $x(s)$ is measurable if and only if it can be approximated uniformly, except possibly on a set of measure zero, by countably-valued functions.*

COROLLARY 1.13. *If $\{x(s)\}$ is a sequence of measurable functions converging weakly a.e. to $x(s)$, then $x(s)$ is measurable.*

For if $\mathfrak{Y}_n$ is a separable sub-B-space containing almost all the values of $x_n(s)$, let $\mathfrak{Y}$ be the separable closed linear hull of a denumerable set dense in $\sum_n \mathfrak{Y}_n$. If $x_n(s) \to x(s)$ weakly at the point s, then ([1], p. 134) some sequence of linear combinations of the elements $\{x_n(s)\}$ converges strongly to $x(s)$ and $x(s)$ is therefore in $\mathfrak{Y}$. Since this statement holds for almost all s, and since $x(s)$ is obviously weakly measurable, the above theorem gives the desired conclusion.

1.14. Suppose that S is euclidean, R_0 is a fixed elementary figure in S, and α is the Lebesgue measure function. If $X(R)$ is defined to $\mathfrak{X}$ from the elementary figures in R_0, then $X(R)$ is *weakly differentiable* at a point s in R_0 if there exists an element $x(s)$ in $\mathfrak{X}$ such that for every f in $\bar{\mathfrak{X}}$

$$f(x(s)) = \lim_{\alpha(I) \to 0} f\left(\frac{X(I)}{\alpha(I)}\right),$$

where I is an arbitrary cube containing s.

THEOREM 1.2. *If $X(R)$ is weakly differentiable a.e. in R_0 to $x(s)$, then $x(s)$ is measurable.*

Let I_0 be a cube containing R_0, and for each n let Π_n be a partition of I_0 into non-overlapping non-degenerate subcubes, with Π_{n+1} a repartition of Π_n and $\lim_{n\to\infty}(\text{norm } \Pi_n)=0$. If s in R_0 is interior to a cube I_{jn} in Π_n, and if $R_0 \supset I_{jn}$, define

$$x_n(s) = \frac{X(I_{jn})}{\alpha(I_{jn})};$$

otherwise set $x_n(s)=\theta$. Each $x_n(s)$ is measurable, and by hypothesis $x_n(s)$ converges weakly to $x(s)$ for almost all s in R_0; from Corollary 1.13 this implies measurability for $x(s)$.

2. **Definition and properties of the integral.** We make the following definition:

DEFINITION 2.1. *A function $x(s)$ from S to $\mathfrak{X}$ is $(\mathfrak{X})$ integrable over measurable E if and only if there exists an x_E in $\mathfrak{X}$ such that*

$$f(x_E) = \int_E f(x(s))d\alpha$$

for every f in $\overline{\mathfrak{X}}$, wherein the integral on the right is the Lebesgue. By definition*

$$\int_E x(s)d\alpha = x_E.$$

A function $x(s)$ is integrable or $(\mathfrak{X})$ integrable if and only if it is $(\mathfrak{X})$ integrable over E for every measurable E.†

2.11. Immediate consequences of the definition are: (1) the integral is single-valued; (2) it is a linear function of the integrand; (3) if $\phi(s)$ is finite for every s and is Lebesgue integrable, and if $x(s)=x_0$, a constant, then $\phi(s)x(s)$ is defined and integrable and $\int_E \phi(s)x(s)d\alpha = x_0 \cdot \int_E \phi(s)d\alpha$ (in particular, $\int_E x_0 d\alpha = x_0 \cdot \alpha(E)$); (4) if $x(s)$ is integrable and $y(s)=x(s)$ a.e., then $y(s)$ is integrable and $\int_E y(s)d\alpha = \int_E x(s)d\alpha$ for every measurable E; (5) if $\mathfrak{X}$ is the space of reals, this definition coincides with the Lebesgue integral.

THEOREM 2.2. *If $x(s)$ from S to $\mathfrak{X}$ is integrable, and if $y=U(x)$ is a linear operation from $\mathfrak{X}$ to $\mathfrak{Y}$, then $y(s)\equiv U(x(s))$ is integrable and $U(\int_E x(s)d\alpha) = \int_E U(x(s))d\alpha$.*

---

* [2], p. 247; [17].

† This definition is analogous to that of I. Gelfand [15]. Gelfand, however, considers functions defined from a linear interval to $\overline{\mathfrak{X}}$.

If any g in $\overline{\mathfrak{Y}}$ is given, the identity $g(y)=g(U(x))=f(x)$ defines an element f of $\overline{\mathfrak{X}}$. By hypothesis the function $f(x(s))=g(y(s))$ is Lebesgue integrable over any measurable E, and

$$f\left(\int_E x(s)d\alpha\right)=\int_E f(x(s))d\alpha=\int_E g(y(s))d\alpha$$

or

$$g\left(U\left(\int_E x(s)d\alpha\right)\right)=\int_E g(y(s))d\alpha.$$

This implies the conclusions of the theorem.

2.21. If (J) is an integral definition for functions $x(s)$ from S to $\mathfrak{X}$ that reduces to the Lebesgue when $\mathfrak{X}$ is the space of reals, and if the (J) integral has the above property of Theorem 2.2, it is clear, on letting $\mathfrak{Y}$ be the real numbers, that every $x(s)$ that is (J) integrable is also integrable and to the same value. Thus the ($\mathfrak{X}$) integral includes both that of Bochner ([9], p. 475) and that of G. Birkhoff ([6], p. 371). From this second inclusion it follows that an ($\mathfrak{X}$) integral is not always of bounded variation. However (Theorems 2.4 and 2.5), it is completely additive and, in the sense of Saks, absolutely continuous as a function of measurable sets.

2.3. If $\sum_n x_n$ is a series of elements of $\mathfrak{X}$, and if every subseries of $\sum_n x_n$ is convergent, then $\sum_n x_n$ is said to be *unconditionally convergent* ([12], p. 33); a function $X(E)$ from Σ to $\mathfrak{X}$ is *completely additive* ([6], p. 365) if and only if for every sequence $\{E_n\}$ of disjoint measurable sets $\sum_n X(E_n)$ is unconditionally convergent and $\sum_n X(E_n)=X(\sum_n E_n)$. A series $\sum_n x_n$ is said to have *property* (O) of Orlicz ([11], p. 244, Theorem 2) if and only if every subseries is weakly convergent to an element of $\mathfrak{X}$.

The next lemma and theorem were proved by Orlicz in [11] (Theorem 2) for the case $\mathfrak{X}$ weakly complete; the general theorem is credited by Banach ([1], p. 240) to the same author without proof or reference. Since we know no published proof to which to refer, and since the lemma is fundamental for our purposes, we include the following demonstration.*

LEMMA 2.31. *If $\sum_n x_n$ has property* (O), *then given $\epsilon>0$ there exists an N_ϵ such that f in $\overline{\mathfrak{X}}$ and $\|f\|=1$ implies $\sum_{n=N_\epsilon}^{\infty}|f(x_n)|<\epsilon$; hence $\|x_n\|\to 0$.*

Let $\mathfrak{Y}$ be the separable closed linear hull of the x_n's. To prove the first part it is sufficient to show that if $\{g_i\}$ is a weakly convergent sequence of functionals over $\mathfrak{Y}$ converging weakly to g_0, then in the space l^1 of absolutely convergent series the norms of the elements $\lambda_i=\{g_i(x_n)-g_0(x_n)\}$ converge to 0. For if there exist an $\epsilon_0>0$, a sequence $\{g_i\}$ in $\overline{\mathfrak{Y}}$, and integers $\{N_i\}$ such

* This lemma has been generalized by Dunford, *Uniformity in linear spaces*, loc. cit.

that $\|g_i\| \leqq 1$, and $N_i \to \infty$, and if

$$\sum_{n=N_i}^{\infty} |\, g_i(x_n)\,| \geqq \epsilon_0 > 0, \qquad \text{for all } i, \tag{1}$$

then since $\mathfrak{Y}$ is separable and $\|g_i\| \leqq 1$, we may suppose ([1], p. 123, Theorem 3) that $\{g_i\}$ is a weakly convergent sequence of functionals over $\mathfrak{Y}$, and hence that there is a g_0 such that $\{g_i\}$ converges weakly to g_0. But $\sum x_n$ has property (O), and for $i = 0, 1, 2, \cdots$, the functional g_i is extensible from $\mathfrak{Y}$ to $\mathfrak{X}$; thus $\sum_{n=1}^{\infty} |\, g_i(x_n)\,| < \infty$, and if $\|\lambda_i\|_{l^1} \to 0$, the inequality (1) is contradicted and the first part of the lemma established.

To prove that $\|\lambda_i\| \to 0$ it is necessary and sufficient that to show ([1], pp. 138–139) $d(\lambda_i) \to 0$ for every linear functional $d(\lambda)$ defined over l^1 for which the associated bounded sequence $\{d_n\}$ is composed only of 0's, $+1$'s, and -1's. If $d(\lambda)$ is such a functional, then by property (O)

$$\begin{aligned} \sum_n d_n f(x_n) &= \sum_{d_n=+1} f(x_n) - \sum_{d_n=-1} f(x_n) \\ &= f(x^+) - f(x^-) \\ &= f(x_0) \end{aligned} \tag{2}$$

for every f in $\bar{\mathfrak{X}}$. Then the sequence $\{\sum_{i=1}^{n} d_i x_i\}$ converges weakly to x_0, so that x_0 must be the strong limit of certain linear combinations of the elements $\{\sum_{i=1}^{n} d_i x_i\}$; hence x_0 is in $\mathfrak{Y}$.

From A and the fact that x_0 is in $\mathfrak{Y}$ it follows that (2) must hold for any g in $\bar{\mathfrak{Y}}$. This implies

$$\begin{aligned} d(\lambda_i) &\equiv \sum_{n=1}^{\infty} d_n \left[g_i(x_n) - g_0(x_n) \right] \\ &= g_i(x_0) - g_0(x_0) \end{aligned}$$

for each i. But $g_i(x) \to g_0(x)$ for every x in $\mathfrak{Y}$ including x_0; hence $d(\lambda_i) \to 0$.

The fact that $\|x_n\| \to 0$ follows immediately on applying B to the result just established.

THEOREM 2.32. *Unconditional convergence is equivalent to property* (O).

Unconditional convergence implies property (O), since convergence to an element implies weak convergence to that element.

If $\sum_n x_n$ is not unconditionally convergent, there is a non-convergent subseries. By inserting parentheses this subseries can be so "grouped" that the resulting series $\sum_m y_m$ has lim sup $\|y_m\| > 0$. If $\sum_n x_n$ has property (O), so does every subseries and every finite grouping of a subseries; then by the lemma $\|y_m\| \to 0$, and we have a contradiction.

THEOREM 2.4. *If $x(s)$ is integrable, then*

$$X(E) = \int_E x(s)d\alpha$$

is a completely additive function of measurable sets.

Let $\{E_n'\}$ be disjoint measurable sets. From the integrability of $x(s)$ and the complete additivity of the Lebesgue integral we have

$$f(X(E')) = \int_{E'} f(x(s))d\alpha = \sum_n \int_{E_n'} f(x(s))d\alpha$$

$$= \sum_n f\left(\int_{E_n'} x(s)d\alpha\right) = \sum_n f(X(E_n')),$$

where $E' = \sum_n E_n'$ and f in $\bar{\mathfrak{X}}$ is arbitrary. If any sequence $\{E_n\}$ of disjoint measurable sets is given, this must hold for every subsequence $\{E_n'\}$; hence $\sum_n X(E_n)$ has property (O) and is therefore unconditionally convergent by Theorem 2.32.

THEOREM 2.5.* *If $x(s)$ is integrable, then*

$$X(E) = \int_E x(s)d\alpha$$

is absolutely continuous.

If not there exist an $\epsilon > 0$ and sequences $\{E_i\}$, $\{f_i\}$ such that $\alpha(E_i) \to 0$, $\|f_i\| = 1$, and

$$\left|\int_{E_i} f_i(x(s))d\alpha\right| = \left|f_i\left(\int_{E_i} x(s)d\alpha\right)\right| = \left\|\int_{E_i} x(s)d\alpha\right\| \geqq \epsilon.$$

Let $\phi(s) = \text{l.u.b.}_i |f_i(x(s))| \leqq \|x(s)\|$, and let $M_n = S[n-1 \leqq \phi(s) < n]$ for each integer $n > 0$. The sets $\{M_n\}$ are disjoint and measurable, $\sum_n M_n = S$, and s in M_n implies $|f_i(x(s))| < n$ for all i.

Since $E_1 = \sum_n E_1 M_n$ and $X(E)$ is completely additive,

$$\epsilon \leqq \left\|\int_{E_1} x(s)d\alpha\right\| = \left\|\sum_{n=1}^{\infty} \int_{E_1 \cdot M_n} x(s)d\alpha\right\|;$$

there must then be an integer $n_1 > n_0 = 0$ such that

$$\left\|\sum_{n=n_0+1}^{n_1} \int_{E_1 \cdot M_n} x(s)d\alpha\right\| > \frac{\epsilon}{2}.$$

* The method of proof here was suggested by E. J. McShane.

Suppose there exist sets $E_{i_1}, \cdots, E_{i_k}$ from $\{E_i\}$ and integers $n_0 < n_1 < \cdots < n_k$ such that

$$\left\| \sum_{n=n_{j-1}+1}^{n_j} \int_{E_{i_j} \cdot M_n} x(s)d\alpha \right\| > \frac{\epsilon}{2}$$

for $j = 1, 2, \cdots, k$. Since $\alpha(E_i) \to 0$ there is an i_{k+1} such that $\alpha(E_{i_{k+1}}) < \epsilon/2n_k$. Then

$$\begin{aligned}
\epsilon &\leqq \left\| \int_{E_{i_{k+1}}} x(s)d\alpha \right\| = \left| \int_{E_{i_{k+1}}} f_{i_{k+1}}(x(s))d\alpha \right| \\
&\leqq \left| \sum_{n=1}^{n_k} \int_{E_{i_{k+1}} \cdot M_n} f_{i_{k+1}}(x(s))d\alpha \right| + \left| \sum_{n=n_{k+1}}^{\infty} \int_{E_{i_{k+1}} \cdot M_n} f_{i_{k+1}}(x(s))d\alpha \right| \\
&\leqq n_k \cdot \alpha(E_{i_{k+1}}) + \left| \sum_{n=n_{k+1}}^{\infty} \int_{E_{i_{k+1}} \cdot M_n} f_{i_{k+1}}(x(s))d\alpha \right| \\
&< \frac{\epsilon}{2} + \left| \sum_{n=n_{k+1}}^{\infty} \int_{E_{i_{k+1}} \cdot M_n} f_{i_{k+1}}(x(s))d\alpha \right|;
\end{aligned}$$

hence there clearly exists an $n_{k+1} > n_k$ such that

$$\left| \sum_{n=n_k+1}^{n_{k+1}} \int_{M_n \cdot E_{i_{k+1}}} f_{i_{k+1}}(x(s))d\alpha \right| > \frac{\epsilon}{2}.$$

Then, if one sets

$$G_{k+1} = \sum_{n=n_k+1}^{n_{k+1}} E_{i_{k+1}} \cdot M_n,$$

it follows that

$$\left\| \int_{G_{k+1}} x(s)d\alpha \right\| \geqq \left| \int_{G_{k+1}} f_{i_{k+1}}(x(s))d\alpha \right| > \frac{\epsilon}{2}.$$

This completes the induction; thus there exist sequences $\{n_k\}$, $\{i_k\}$ of integers such that $n_k < n_{k+1}$, and if G_{k+1} is defined as above, for $k = 0, 1, 2, \cdots$,

$$\left\| \int_{G_{k+1}} x(s)d\alpha \right\| > \frac{\epsilon}{2}.$$

But the sets G_k are disjoint and measurable, so that $\sum_{k=1}^{\infty} \int_{G_k} x(s)d\alpha = \int_{\sum G_k} x(s)d\alpha$, whence $\| \int_{G_k} x(s)d\alpha \| \to 0$, a contradiction.

COROLLARY 2.51. *If $x(s)$ is integrable, then for any bounded set $\mathfrak{M} = [f]$ of elements of $\bar{\mathfrak{x}}$ the integrals $\int_E |f(x(s))|\, d\alpha$ are equi-absolutely continuous.*

If $\|f\| \leqq K$, we have $\left|\int_E f(x(s))d\alpha\right| = \left|f\left(\int_E x(s)d\alpha\right)\right| \leqq \|f\| \cdot \left\|\int_E x(s)d\alpha\right\| \leqq K \cdot \left\|\int_E x(s)d\alpha\right\| < \epsilon$ for $\alpha(E)$ less than a suitably chosen δ. Thus if $E_f^+ \equiv E[f(x(s)) \geqq 0]$ and $E_f^- = E - E_f^+$, we have

$$\int_E |f(x(s))|\, d\alpha = \left|\int_{E_f^+} f(x(s))d\alpha\right| + \left|\int_{E_f^-} f(x(s))d\alpha\right| < 2\epsilon$$

for $\alpha(E) < \delta$, since $\alpha(E)$ is non-negative and $E_f^+ + E_f^- = E$.

3. **Operations defined by integrable functions.** In this section S will be the closed interval $[0, 1]$ and $\alpha(E)$ the Lebesgue measure function. According to custom L^p, $(1 \leqq p < \infty)$, will denote the class of real-valued measurable functions $\phi(s)$ defined over S and having $|\phi(s)|^p$ (L) integrable, and L^∞ the space of the real-valued essentially bounded and measurable functions defined over S. It is well known that L^p forms a B-space when $\|\phi\| = \left(\int_0^1 |\phi(s)|^p ds\right)^{1/p}$ for $1 \leqq p < \infty$ and $\|\phi\| = \text{ess. sup.}_S |\phi(s)|$ for $p = \infty$. There is no loss of generality in supposing that every element $\phi(:)$ of L^p has no infinite functional values.

DEFINITION 3.1. *If $1 \leqq p' \leqq \infty$, then $\mathfrak{L}^{p'}(\mathfrak{X})$ is, for a given $\mathfrak{X}$, the class of all functions $x(s)$ from S to $\mathfrak{X}$ having $f(x(:))$ in $L^{p'}$ for every f in $\overline{\mathfrak{X}}$;* if $p' = \infty$, the added condition*

$$\text{l.u.b.}_{\|f\|=1} \left(\text{ess.sup.}_S |f(x(s))|\right) < \infty$$

must be satisfied by each $x(:)$ in $\mathfrak{L}^\infty(\mathfrak{X})$; this l.u.b. *is denoted by $((x))_\infty$. The class $\mathfrak{L}_0^{p'}(\mathfrak{X})$ is defined to be the subclass of $\mathfrak{L}^{p'}(\mathfrak{X})$ composed of integrable functions.*

3.11. The class $\mathfrak{L}^\infty(\mathfrak{X})$ always includes the weakly measurable and essentially bounded functions, and $((x))_\infty \leqq \text{ess. sup.}_S \|x(s)\|$. A partial converse is that if the separably valued function $x(:)$ is in $\mathfrak{L}^\infty(\mathfrak{X})$, then $x(s)$ is not only measurable (by Theorem 1.1) but also essentially bounded. Suppose $x(s)$ is in $\mathfrak{Y}$, a separable subspace, for almost all s, and that $x(:)$ is in $\mathfrak{L}^\infty(\mathfrak{X})$. Let $\{g_i\}$ be weakly dense in the surface of the unit sphere of $\overline{\mathfrak{Y}}$, and let A be the set of values of s such that $\|x(s)\| > ((x))_\infty$ and A_i the set such that $|g_i(x(s))| > ((x))_\infty$. If $A' = S - A$, $A_i' = S - A_i$, then by the proof of Theorem 1.1 we have $A' = \prod_i A_i'$ and hence $A = \sum A_i$. Since $\alpha(A_i) = 0$, this implies that $x(s)$ is essentially bounded and that ess. sup.$_S \|x(s)\| = ((x))_\infty$.

Thus if $\mathfrak{X}$ is separable, $\mathfrak{L}^\infty(\mathfrak{X})$ is the class $S_\infty(\mathfrak{X})$ ([9], p. 474) composed of essentially bounded and measurable functions, that is, essentially bounded Bochner integrable functions. Since $\mathfrak{L}^\infty(\mathfrak{X}) \supset \mathfrak{L}_0^\infty(\mathfrak{X}) \supset S_\infty(\mathfrak{X})$, we have, for separable $\mathfrak{X}$,

* This is due to Dunford (cf. [8]) and constitutes a direct generalization of the $(\mathfrak{X})$ integral.

$$S_\infty(\mathfrak{X}) = \mathfrak{L}^\infty(\mathfrak{X}) = \mathfrak{L}_0^\infty(\mathfrak{X}).$$

THEOREM 3.2. *If $x(:)$ is in $\mathfrak{L}^{p'}(\mathfrak{X})$, $(1 \leqq p' \leqq \infty)$, the operation*

$$U(f) = f(x(:))$$

from $\overline{\mathfrak{X}}$ to $L^{p'}$ is linear, and

$$\|U\| = \text{l.u.b.}_{\|f\|=1} \left(\int_0^1 |f(x(s))|^{p'} ds \right)^{1/p'}, \qquad 1 \leqq p' < \infty,$$

$$\|U\| = ((x))_\infty, \qquad p' = \infty.$$

The additivity of U is trivial. In case $p' = \infty$

$$\|U(f)\| = \text{ess.sup.}_S |f(x(s))| \leqq \|f\| \cdot ((x))_\infty,$$

and U is linear. If $p' < \infty$ and $f_n \to f_0$ in $\overline{\mathfrak{X}}$, then $f_n(x(s)) \to f_0(x(s))$ for every s, and by Fatou's lemma

$$\liminf_n \int_0^1 |f_n(x(s))|^{p'} ds \geqq \int_0^1 |f_0(x(s))|^{p'} ds$$

or $\liminf_n \|U(f_n)\| \geqq \|U(f_0)\|$, and U is again linear. The remainder of the theorem follows from the fact that $\|V\| = \text{l.u.b.}_{\|y\|=1} \|V(y)\|$ for an arbitrary linear operation $V(y)$.

DEFINITION 3.21. *In the linear space $\mathfrak{L}^{p'}(\mathfrak{X})$ the norm of $x(:)$ is $((x))_{p'} = \|U\|$.*

3.22. In the linear normed subspace $\mathfrak{L}_0^1(\mathfrak{X})$ a sequence $\{x_n\}$ is a Cauchy sequence if and only if $\text{l.u.b.}_E \|\int_E (x_m - x_n) ds\| \to 0$.* For

$$\text{l.u.b.}_E \left\| \int_E (x_m - x_n) ds \right\| = \text{l.u.b.}_E \ \text{l.u.b.}_{\|f\|=1} \left| \int_E f(x_m(s) - x_n(s)) ds \right|$$

$$\leqq \text{l.u.b.}_{\|f\|=1} \int_0^1 |f(x_m - x_n)| \, ds$$

$$\leqq ((x_m - x_n))_1;$$

and, on the other hand, if $\text{l.u.b.}_E \|\int_E (x_m - x_n) ds\| < \epsilon$, then for $\|f\| = 1$

$$\int_0^1 |f(x_m - x_n)| \, ds = \left| \int_{E_f^+} f(x_m - x_n) ds \right| + \left| \int_{E_f^-} f(x_m - x_n) ds \right|$$

$$\leqq \left\| \int_{E_f^+} (x_m - x_n) ds \right\| + \left\| \int_{E_f^-} (x_m - x_n) ds \right\| < 2\epsilon.$$

* See [6], p. 368, where this l.u.b. is used to norm the Birkhoff integrable functions.

Thus in $\mathfrak{L}_0^1(\mathfrak{X})$ these two definitions of norm are topologically equivalent; they are not, however, identical.†

In passing we note the obvious facts that if $x(:)$ is in $\mathfrak{L}^{p'}(\mathfrak{X})$, then $x(:)$ is in $\mathfrak{L}^{q'}(\mathfrak{X})$ for $1 \leqq q' \leqq p'$, and that

$$\lim_{m,n\to\infty} ((x_m - x_n))_{p'} = 0 \to \lim_{m,n} ((x_m - x_n))_{q'} = 0 \to \lim_{m,n} ((x_m - x_n))_1 = 0$$

$$\to \lim_{m,n} \text{l.u.b.}_E \left\| \int_E (x_m - x_n) \right\| = 0$$

$$\to \lim_{m,n} \left\| \int_E (x_m - x_n) \right\| = 0$$

for every measurable E, the arrows meaning "implies that."

In the remainder of this section we shall consider pairs p, p' of "conjugate exponents," that is, numbers p, p' connected by the relations $1/p+1/p'=1$ for $1<p<\infty$ (or $1<p'<\infty$), $p'=\infty$ for $p=1$, and $p'=1$ for $p=\infty$, and shall consider the operations from L^p to $\mathfrak{X}$ defined by elements of $\mathfrak{L}_0^{p'}(\mathfrak{X})$. Given L^p we shall deal only with $\mathfrak{L}^{p'}(\mathfrak{X})$, where p and p' are conjugate, and given $\mathfrak{L}^q(\mathfrak{X})$, only with $L^{q'}$, q and q' conjugate.

THEOREM 3.3. *If $x(:)$ is in $\mathfrak{L}^{p'}(\mathfrak{X})$, then for a fixed $\phi(:)$ in L^p the integral*

$$F_\phi(f) = \int_0^1 f(x(s))\phi(s)ds \tag{1}$$

defines a linear functional F_ϕ over $\overline{\mathfrak{X}}$,‡ and the operation

$$F_\phi = U^*(\phi) \tag{2}$$

so defined from L^p to $\overline{\overline{\mathfrak{X}}}$ is linear, with $\|U^*\| = ((x))_{p'}$.

For a fixed $x(:)$ in $\mathfrak{L}^{p'}(\mathfrak{X})$, $(1 \leqq p' \leqq \infty)$, the operation $U(f)=f(x(:))$ from $\overline{\mathfrak{X}}$ to $L^{p'}$ is linear by Theorem 3.2. Then if g_ϕ is the linear functional over $L^{p'}$ generated by the element $\phi(:)$ of L^p, it follows that

$$g_\phi(U(f)) = \int_0^1 f(x(s))\phi(s)ds = F_\phi(f) \tag{3}$$

defines a linear functional $F_\phi(f)$ over $\overline{\mathfrak{X}}$, that is, an element F_ϕ of $\overline{\overline{\mathfrak{X}}}$. If $p'<\infty$, so that g is a linear functional over $L^{p'}$ only if it is so generated, the operation

$$F_\phi = U^*(\phi)$$

† If we take $n^2=2$ in example 4 of [6], then $((x))_1 \geqq 2^{-1/2}$ while $\text{l.u.b.}_E \| \int_E x(s)ds \| < 2^{-1/2}$.

‡ See [8].

thus defined from $\overline{L^{p'}} = L^p$ to $\mathfrak{X}$ is by definition the operation $\overline{U}$ adjoint to U; hence U^* is linear and $\|U^*\| = \|\overline{U}\| = \|U\| = ((x))_{p'}$.

If $p' = \infty$, then $\overline{U}$ is from $\overline{L^\infty}$ to $\overline{\mathfrak{X}}$, where $\overline{L^\infty}$ is not an L^p class ([13], p. 875; [16]). But $\overline{L^\infty}$ includes L^1 in the sense that any $\phi(:)$ in L^1 defines a linear functional

$$g_\phi(\phi') = \int_0^1 \phi'(s)\phi(s)ds, \qquad \phi' \text{ in } L^\infty,$$

over L^∞, where $\|g_\phi\|$ is $\int_0^1 |\phi(s)|\,ds$, the norm of $\phi(:)$ as an element of L^1. Thus $U^*(\phi)$ maps this imbedded subspace onto $\overline{\mathfrak{X}}$, and by (2) and equation (3) it is evident that U^* coincides with $\overline{U}$ where U^* is defined. Hence U^* is linear and $\|U^*\| \leqq ((x))_\infty$. If $\|U^*\| < ((x))_\infty$, then there is an f with $\|f\| = 1$ and $\|U^*\| <$ ess. sup.$_S$ $|f(x(s))|$, that is, $\|U^*\| < |f(x(s))|$ on a set of measure greater than zero. Hence there is an f_0 (either $+f$ or $-f$) and a set S_0 such that $\|f_0\| = 1$, $\alpha(S_0) > 0$, and $f_0(x(s)) > \|U^*\| \geqq 0$ on S_0. Letting $\phi_0 = 1/\alpha(S_0)$ times the characteristic function of S_0, we have

$$\begin{aligned}\|U^*\| &= \|U^*\| \cdot \|\phi_0\| \geqq \|U^*(\phi_0)\| = \|F_{\phi_0}\| \geqq |F_{\phi_0}(f_0)| \\ &= \left| \int_{S_0} f_0(x(s)) \cdot \frac{1}{\alpha(S_0)}\, ds \right| > \int_{S_0} \|U^*\| \cdot \frac{1}{\alpha(S_0)}\, ds = \|U^*\|,\end{aligned}$$

which is impossible. Thus $\|U^*\| = ((x))_\infty = \|U\| = \|\overline{U}\|$.

THEOREM 3.4. *In order that $x(:)$ be in $\mathfrak{L}_0^{p'}(\mathfrak{X})$, it is necessary and sufficient that $x(:)\phi(:)$ be integrable for every ϕ in L^p. If $x(:)$ is in $L_0^{p'}(\mathfrak{X})$, then the operation*

$$V(\phi) = \int_0^1 x(s)\phi(s)ds \tag{4}$$

from L^p to $\mathfrak{X}$ is linear, with $\|V\| = ((x))_{p'}$. Moreover, the operation $\overline{V}(f)$ adjoint to V is from $\overline{\mathfrak{X}}$ to $L^{p'}$ and assigns to each f in $\overline{\mathfrak{X}}$ the linear functional over L^p generated by the element $U(f) = f(x(:))$ of $L^{p'}$.

Suppose that $x(:)$ is in $\mathfrak{L}_0^{p'}(\mathfrak{X})$. If $\phi(s)$ is a step-function, the integrability of $x(s)\phi(s)$ follows from the linearity and set-additivity of the integral. If $\phi(:)$ in L^p is arbitrary, let $\{\phi_n\}$ be step-functions converging to ϕ in L^p. For any measurable E and any m, n there exists, by B, an f_{mn} in $\overline{\mathfrak{X}}$ such that $\|f_{mn}\| = 1$ and

$$\left\| \int_E x(s)\phi_m(s) - \int_E x(s)\phi_n(s) \right\| = \left| f_{mn}\left(\int_E (\phi_m - \phi_n)x(s)ds \right) \right|;$$

on applying Hölder's inequality, we obtain

$$\left\| \int_E x\phi_m ds - \int_E x\phi_n ds \right\| \leqq \int_E | f_{mn}((\phi_m - \phi_n)x) | \, ds$$
$$\leqq \|f_{mn}(x(:))\|_{L^{p'}} \|\phi_m - \phi_n\|_{L^p}$$
$$\leqq ((x))_{p'} \|\phi_m - \phi_n\|_{L^p} \to 0.$$

If $x_E \equiv \lim_n \int_E x(s)\phi_n(s)ds$, then for any f in $\bar{\mathfrak{X}}$ it follows that

$$f(x_E) = \lim_n f\left(\int_E x\phi_n ds \right) = \lim_n \int_E f(x(s)) \cdot \phi_n(s) ds$$
$$= \int_E f(x(s)) \cdot \phi(s) ds = \int_E f(x(s) \cdot \phi(s)) ds,$$

and $x(s) \cdot \phi(s)$ is integrable.

If $x(s)$ is integrable for every ϕ in L^p, then $f(x(s))\phi(s)$ is summable for every ϕ; hence $f(x(:))$ is in $L^{p'}$, that is, $x(:)$ is in $\mathfrak{L}^{p'}(\mathfrak{X})$. The integrability of $x(s)$ results from letting ϕ be the characteristic function of measurable E.

If x_0 in $\mathfrak{X}$ is fixed, the identity

$$F_0(f) \equiv f(x_0), \qquad \text{for all } f \text{ in } \bar{\mathfrak{X}} \tag{5}$$

defines a linear functional F_0 over $\bar{\mathfrak{X}}$. If one sets

$$F_0 = W(x_0),$$

the operation W is linear and norm-preserving from $\mathfrak{X}$ to $\bar{\bar{\mathfrak{X}}}$ and is 1-1 from $\mathfrak{X}$ to its contradomain, the subspace $\bar{\bar{\mathfrak{X}}}'$ of $\bar{\bar{\mathfrak{X}}}$ composed of all elements having a representation of the form (5) ([1], p. 189, proof of Theorem 13). Since W is 1-1 and norm-preserving, and since $\mathfrak{X}$ is a B-space, $\bar{\bar{\mathfrak{X}}}'$ is also a B-space and

$$x = W^{-1}(F)$$

from $\bar{\bar{\mathfrak{X}}}'$ to $\mathfrak{X}$ is linear and norm-preserving.

For ϕ in L^p the integrability of $x(s)\phi(s)$ implies

$$F_\phi(f) = \int_0^1 f(x(s))\phi(s)ds = f\left(\int_0^1 x(s)\phi(s)ds \right),$$

and F_ϕ is in $\bar{\bar{\mathfrak{X}}}'$, with

$$W^{-1}(F_\phi) = \int_0^1 x(s)\phi(s)ds$$

or

$$W^{-1}(U^*(\phi)) = V(\phi).$$

The operation $V(\phi)$ must then be linear, and, since W^{-1} is norm-preserving,

$$\|V\| = \|U^*\| = ((x))_{p'}.$$

To complete the proof suppose f in $\bar{\mathfrak{X}}$, and let f' be the linear functional over L^p defined by $f'(\phi) \equiv f(V(\phi))$; that is, let $f' = \bar{V}(f)$ where $\bar{V}$ is adjoint to V. Then

$$f'(\phi) = f\left(\int_0^1 x(s)\phi(s)ds\right) = \int_0^1 f(x(s)\cdot\phi(s))ds$$

holds for every ϕ in L^p, where $f(x(:))$ is in $L^{p'}$; thus f' is the linear functional over L^p generated by the element $f(x(:)) = U(f)$ of $L^{p'}$.

COROLLARY 3.41. *If $x(s)$ is integrable and $\phi(s)$ is a measurable and essentially bounded real-valued function, then $x(s)\phi(s)$ is integrable.*

4. **Convergence and approximation of integrable functions.** We begin with the following theorem:

THEOREM 4.1. *If $\{x_n(s)\}$ is a sequence of integrable functions converging weakly in measure to $x(s)$, and if $\lim_n \int_E x_n(s)d\alpha$ exists for every measurable E, then $x(s)$ is integrable and $x_E \equiv \lim_n \int_E x_n(s)d\alpha = \int_E x(s)d\alpha$.*

Since $f(x_n(s)) \to f(x(s))$ in measure and since also $f(x_E) = \lim_n f(\int_E x_n(s)d\alpha) = \lim_n \int_E f(x_n(s))d\alpha$, from real-function theory it follows that $f(x(s))$ is integrable and that

$$f(x_E) = \lim_n \int_E f(x_n(s))d\alpha = \int_E f(x(s))d\alpha$$

for any f in $\bar{\mathfrak{X}}$ and any measurable E. Thus $x(s)$ is integrable and $\int_E x(s)d\alpha = x_E = \lim_n \int_E x_n(s)d\alpha$.

THEOREM 4.2. *If $\{x_n\}$ is a Cauchy sequence in $\mathfrak{L}_0^p(\mathfrak{X})$ and $x_n(s) \to x(s)$ weakly in measure, then $x(:)$ is in $\mathfrak{L}_0^p(\mathfrak{X})$ and $((x - x_n))_p \to 0$.*

Since $\{x_n\}$ is a Cauchy sequence,

$$\lim_{m,n\to\infty} \text{l.u.b.}_{\|f\|=1} (\|f(x_m(:)) - f(x_n(:))\|_{L^p}) = 0;$$

it follows for any f, on considering $f_u = f/\|f\|$, that the functions $f(x_n(:))$ converge as elements in L^p. Since $f(x_n(s)) \to f(x(s))$ in measure, the conclusion is that $f(x(:))$ is in L^p, that is, $x(:)$ is in $\mathfrak{L}^p(\mathfrak{X})$. On the other hand, the hypotheses here imply those of Theorem 4.1, so that $x(:)$ is in $\mathfrak{L}_0^p(\mathfrak{X})$.

Now by hypothesis, given an $\epsilon < 0$ there exists an N_ϵ such that if $m, n \geqq N_\epsilon$, then

$$\text{l.u.b.}_{\|f\|=1} (\|f(x_m(:)) - f(x_n(:))\|_{L^p}) < \epsilon;$$

thus for any fixed $m \geqq N_\epsilon$ and for any f with $\|f\| = 1$ we obtain

$$\|f(x_m(:)) - f(x(:))\|_{L^p} \leqq \epsilon$$

on letting $n \to \infty$ and recalling that $f(x_n(:)) \to f(x(:))$ in L^p and that in L^p the norm is a continuous function. This last inequality holds for any $\|f\| = 1$ and any $m \geqq N_\epsilon$; then $((x_m - x))_p \leqq \epsilon$ for $m \geqq N_\epsilon$, or $((x_m - x))_p \to 0$.

THEOREM 4.3. *The function $x(s)$ is integrable and measurable if and only if there exist step-functions $\{y_n(s)\}$ such that $\{y_n(s)\} \to x(s)$ a.e. and $((y_n - x))_1 \to 0$.*

That the existence of such a sequence implies $x(s)$ measurable and integrable is evident from Theorem 4.2. In the proof of the converse two lemmas will be used.

LEMMA 4.31. *The theorem is true for integrable countably-valued functions.*

Let $x(s)$ have the constant value x_i on E_i, the sets $\{E_i\}$ being countable, measurable, and disjoint, and let $S = \sum E_i$. Since $\sum_i \int_{E_i} x(s) d\alpha$ converges unconditionally, there exists (Theorem 2.32 and Lemma 2.31) for each $1/2^j$ an N_j such that $\|f\| \leqq 1$ implies

$$\sum_{N_j+1}^{\infty} |f(x_i)| \alpha(E_i) = \sum_{N_j+1}^{\infty} \left| f\left(\int_{E_i} x(s) d\alpha \right) \right| < 1/2^j,$$

with $N_1 < N_2 < \cdots$. By defining $y_n(s) = x_i$ on E_i for $i \leqq N_n$ and $y_n(s) = \theta$ elsewhere, one obtains $\{y_n(s)\} \to x(s)$ everywhere and

$$\int_S |f(x(s) - y_n(s))| \, d\alpha = \sum_{N_n+1}^{\infty} \int_{E_i} |f(x(s))| \, d\alpha < 1/2^n$$

for $\|f\| \leqq 1$; hence $((x - y_n))_1 \to 0$.

LEMMA 4.32. *If $x(s)$ is measurable and integrable, it can be uniformly approximated, except possibly on a set of measure zero, by integrable countably-valued functions.*

In virtue of Corollary 1.12 there are countably-valued functions $x_n(s)$ such that $\|x(s) - x_n(s)\| < 1/n$ uniformly for s in S_0, where $\alpha(S_0) = \alpha(S)$. The measurable function $x - x_n$ is then essentially bounded, therefore Bochner integrable, and hence ($\mathfrak{X}$) integrable, which implies integrability for $x_n = x - (x - x_n)$.

Returning to the proof of the theorem, we suppose $x(s)$ integrable and measurable and $x_n(s)$ a countably-valued integrable function within $1/n$ of $x(s)$ essentially uniformly (Lemma 4.32). Then for $\|f\| \leqq 1$ we have

$$\int_S |f(x - x_n)| \, d\alpha \leqq \int_S \|x - x_n\| d\alpha < \alpha(S) \cdot 1/n,$$

so that $((x - x_n))_1 < \alpha(S) \cdot 1/n$.

If n is fixed, the integrable c.v. function $x_n(s)$ is a.e. the limit of a sequence of step-functions, which therefore converge almost uniformly ([5], p. 442) to $x_n(s)$. Hence there exists a step-function $y_n(s)$ such that $\|x_n(s)-y_n(s)\|<1/n$ except on a set of measure at most $1/2^n$, while (from the second conclusion of the first lemma)

$$((x_n - y_n))_1 < \alpha(S)\cdot 1/n.$$

Collecting inequalities, we have $\|x(s)-y_n(s)\|<2/n$ except on a set of measure at most $1/2^n$, and $((x-y_n))_1<\alpha(S)\cdot 2/n$. It follows that $y_n(s)\to x(s)$ almost uniformly, and therefore a.e., and that $((x-y_n))_1\to 0$.

COROLLARY 4.33. *If $\mathfrak{X}$ is separable, then $\mathfrak{L}_0^1(\mathfrak{X})$ is separable.*

5. **The ($\mathfrak{X}$) integral as related to others.** In addition to the Bochner (Bn), Birkhoff (Bk), and Gelfand extensions of the Lebesgue integral, and Dunford's first integral ([5]; this is equivalent to Bochner's, cf. [9], p. 475), two other definitions, also due to Dunford, have been given ([7], [8]). The last two integrals will be here denoted as the (D1) integral and the (D2) integral, respectively. The first, a direct generalization of the Bochner integral, is defined as follows: $x(s)$ is (D1) integrable if and only if there exist step-functions $x_n(s)$ such that $x_n(s)\to x(s)$ a.e. and $\lim_n \int_E x_n(s)d\alpha$ exists for every measurable E. This limit is (D1) $\int_E x(s)d\alpha$. The second, as we have noted earlier, is a generalization of the ($\mathfrak{X}$) integral; it requires merely that $x(:)$ be in $\mathfrak{L}^1(\mathfrak{X})$ and defines (D2) $\int_E x(s)d\alpha$, to be the element F_E of $\overline{\mathfrak{X}}$ given by $F_E(f)=\int_E f(x(s))d\alpha$.

In paragraph 2.21 we noted that the ($\mathfrak{X}$) integral includes both the (Bn) and the (Bk). The connection between the (D1) and (D2) integrals and the ($\mathfrak{X}$) integral will now be considered.

THEOREM 5.1. *A function $x(s)$ is measurable and integrable if and only if $x(s)$ is* (D1) *integrable. The integral of $x(s)$ has the same value for the two definitions.*

This theorem is seen to follow immediately from Theorem 4.3 and Theorem 4.1.

COROLLARY 5.11. *For measurable functions the ($\mathfrak{X}$), (Bk), and (D1) integrals are equivalent.*

In view of the theorem and the fact that (Bk) integrability implies ($\mathfrak{X}$) integrability to the same value, it is sufficient to show that a measurable integrable $x(s)$ is (Bk) integrable; that is, ([6], Theorem 13), given $\epsilon>0$ there exist disjoint measurable sets $\{E_i\}$, $i=0,1,\cdots,n,\cdots$, such that $\sum_{i=0}^{\infty}E_i=S$, the series $\sum_{i=0}^{\infty}\alpha(E_i)x(s_i)$ is unconditionally convergent for any choice of s_i in

E_i, and s_i, s_i' in E_i implies $\|\sum_1^\infty \alpha(E_i)x(s_i) - \sum_1^\infty \alpha(E_i)x(s_i')\| < \epsilon$. From Theorem 4.32 there exist a c.v. integrable $y(s)$ and a set E_0 such that $\alpha(E_0) = 0$

$$\|x(s) - y(s)\| < \epsilon/4\alpha(S), \qquad \text{for } s \text{ in } S_0 = S - E_0, \tag{1}$$

and

$$\left\| \int_{S_0} x(s)d\alpha - \int_{S_0} y(s)d\alpha \right\| < \epsilon/4. \tag{2}$$

Let $\{E_i\}$, $(i = 1, 2, \cdots)$, be disjoint measurable sets with $\sum_1^\infty E_i = S_0$ and $y(s) = y_i$ on E_i. Then $\int_{S_0} y(s)d\alpha = \sum_1^\infty \alpha(E_i)y_i$ is unconditionally convergent; hence, since (1) implies that for arbitrary s_i in E_i, $(i = 0, 1, \cdots)$, the inequality $\sum_1^\infty \|\alpha(E_i)(x(s_i) - y_i)\| < \epsilon/4$ holds, the series

$$\sum_0^\infty \alpha(E_i)x(s_i) = \sum_1^\infty \alpha(E_i)x(s_i) = \sum_1^\infty \alpha(E_i)(x(s_i) - y_i) + \sum_1^\infty \alpha(E_i)y_i$$

is unconditionally convergent, being the sum of two such series. Moreover, $\|\sum_0^\infty \alpha(E_i)x(s_i) - \sum_1^\infty \alpha(E_i)y_i\| < \epsilon/4$, and this fact combined with (2) gives $\|\sum_0^\infty \alpha(E_i)x(s_i) - \int_{S_0} x(s)d\alpha\| < \epsilon/2$, which completes the proof.

A consequence of Corollary 5.11 is the failure of the Fubini theorem to hold for the (D1), ($\mathfrak{X}$), and (D2) integrals, since ([6], example 6) the theorem fails for the (Bk) integral when $\mathfrak{X}$ is the separable space L^2.

THEOREM 5.2. *The integrals of two measurable and integrable functions $x(s)$, $y(s)$ coincide over every measurable set if and only if $x(s) = y(s)$ a.e.*

If the integrals of measurable $x(s)$ and $y(s)$ coincide everywhere, then $z(s) = x(s) - y(s)$ is separably-valued and has its integral vanishing identically. Let S_0 be a subset in S such that $\alpha(S - S_0) = 0$ and $z(S_0) \subset \mathfrak{Y}$, where $\mathfrak{Y}$ is a separable subspace of $\mathfrak{X}$. Letting $\{g_i\}$ be a sequence weakly dense in the surface of the unit sphere of $\bar{\mathfrak{Y}}$, we find as in paragraph 3.11 that $A = \sum A_i$, where $A = S_0[\|z(s)\| > 0]$ and $A_i = S_0[|g_i(z(s))| > 0]$. Since $\int_E z(s)d\alpha = \theta$ for all E, it results that $\int_E g_i(z(s))d\alpha \equiv 0$ in E for each i, and therefore that $|g_i(z(s))| = 0$ a.e. in S_0. Thus $\alpha(A_i) = 0$ which implies $\alpha(A) = 0$, so that $z(s) = \theta$ a.e. in S_0 and hence a.e. in S.*

The remaining part of the theorem follows from (4) of paragraph 2.11.

THEOREM 5.3. *If measurable $x(s)$ is in $\mathfrak{L}^1(\mathfrak{X})$, then $x(s)$ is integrable if and only if $F_E = (\mathrm{D2})\int_E x(s)d\alpha$ is absolutely continuous.*

---

* Theorem 5.2 can be regarded, in the light of 5.11, as a corollary of Theorem 24 of [6]. In the proofs of Corollary 5.11 and Theorem 5.2 we have essentially repeated Birkhoff's arguments for Theorems 22 and 24 of [6].

If $x(s)$ is in $\mathfrak{L}_0^1(\mathfrak{X})$, then

$$F_E(f) = \int_E f(x(s))d\alpha = f\left(\int_E x(s)d\alpha\right)$$

for all f; therefore $\|F_E\| = \|\int_E x(s)d\alpha\|$ where the first norm is taken over $\bar{\mathfrak{X}}$ and the second over $\mathfrak{X}$. The absolute continuity of F_E now follows from Theorem 2.5.

Suppose $x(s)$ measurable and in $\mathfrak{L}^1(\mathfrak{X})$. Let $\mathfrak{Y}$ be a separable subspace containing almost all of the values $x(s)$; then $x(:)$ is in $\mathfrak{L}^1(\mathfrak{Y})$, since any linear functional over $\mathfrak{Y}$ is extensible to $\mathfrak{X}$. To show that $x(:)$ is in $\mathfrak{L}_0^1(\mathfrak{Y})$, and therefore that $x(:)$ is in $\mathfrak{L}_0^1(\mathfrak{X})$, it is sufficient, $\mathfrak{Y}$ being separable, to prove that for any measurable E the linear functional $F_E(g)$ over $\bar{\mathfrak{Y}}$ is weakly continuous ([1], p. 131). If $g_n(y)\to g_0(y)$ for every y in $\mathfrak{Y}$, then $\|g_n\| \leqq K$, $(n=0, 1, 2, \cdots)$; and if F_E is a.c., then

$$\left|\int_E g_n(x(s))d\alpha\right| = |F_E(g_n)| \leqq \|g_n\| \cdot \|F_E\| \leqq K\epsilon$$

for $\alpha(E)<\delta$. The integrals $\int_E g_n(x(s))d\alpha$ are thus equi-absolutely continuous; since $g_n(x(s))\to g_0(x(s))$ for almost all s, it follows that

$$F_E(g_n) = \int_E g_n(x(s))d\alpha \to \int_E g_0(x(s))d\alpha = F_E(g_0)$$

for every E. Thus for any fixed E the functional $F_E(g)$ over $\bar{\mathfrak{Y}}$ is weakly continuous, so that a y_E in $\mathfrak{Y}$ exists such that $g(y_E)=F_E(g)$ for all g in $\bar{\mathfrak{Y}}$. Since any linear functional over $\mathfrak{X}$ defines one over $\mathfrak{Y}$, and since y_E is in $\mathfrak{Y}$, it follows that $F_E(f)=f(y_E)$ for all f in $\bar{\mathfrak{X}}$, and $x(s)$ is integrable.

COROLLARY 5.31. *If measurable $x(s)$ is in $\mathfrak{L}^p(\mathfrak{X})$, $(p>1)$, then $x(:)$ is in $\mathfrak{L}_0^p(\mathfrak{X})$. If $\mathfrak{X}$ is weakly complete, this is also true for $p=1$.*

Suppose $p>1$. If

$$F_E(f) \equiv \int_E f(x(s))d\alpha = \int_0^1 f(x(s))\phi_E(s)ds,$$

where $\phi_E(s)$ is the characteristic function of E, then

$$|F_E(f)| \leqq \|f(x(:))\|_{L^p} \cdot \|\phi_E(:)\|_{L^{p'}} \leqq ((x))_p \cdot |\alpha(E)|^{1/p'} < \epsilon$$

if $\|f\| \leqq 1$ and $\alpha(E)<\epsilon^{p'}/[((x))_p]^{p'}$. Thus $\|F_E\| \leqq \epsilon$ when $\alpha(E)$ is sufficiently small; hence $x(:)$ is integrable by the theorem.

Suppose $\mathfrak{X}$ weakly complete and $p=1$. From Corollary 1.12 and Theorem 4.2 it clearly suffices to consider c.v. functions. For such an $x(s)$

$$\infty > \int_E |f(x(s))|\, d\alpha = \sum |f(x_i)| \cdot \alpha(E_i \cdot E),$$

so that $\sum x_i \alpha(E_i \cdot E)$ is a weakly convergent series in the weakly complete space $\mathfrak{X}$. Then there exists an x_E such that

$$f(x_E) = \sum f(x_i) \cdot \alpha(E_i \cdot E) = \int_E f(x(s)) d\alpha,$$

and $x(s)$ is integrable.

COROLLARY 5.32. *If* $1 \leqq p \leqq \infty$ *and* $1 \leqq q < \infty$, *then* $\mathfrak{L}^p(L^q) = \mathfrak{L}_0^p(L^q)$ *and* $\mathfrak{L}^p(l^q) = \mathfrak{L}_0^p(l^q)$.

6. **Concerning completely continuous operations.** We prove the following lemma:

LEMMA 6.11. *If* $x(s)$ *is a step-function from* $[0, 1]$ *to* $\mathfrak{X}$, *the operations*

$$U(f) = f(x(:))$$

$$V(\phi) = \int_0^1 x(s)\phi(s)ds$$

from $\overline{\mathfrak{X}}$ *to* $L^{p'}$ *and from* L^p *to* $\mathfrak{X}$ *are c.c.* (*completely continuous*).

If U is c.c., so are $\overline{U}$ and (Theorem 3.3) U^*; since $V(\phi) = W^{-1}(U^*(\phi))$, where W^{-1} is linear, this implies the same property for V. Suppose $x(s) = x_i$ on E_i, $(\sum_1^N E_i = [0, 1])$; then $U(f)$ is the function having the constant value $f(x_i)$ on E_i. If $\|f_n\| \leqq K$, so that $|f_n(x_i)| \leqq K \cdot \|x_i\|$, then clearly there exists a subsequence $\{f_n'\}$ such that the functions $\{f_n'(x(s))\}$ converge uniformly over $[0, 1]$ and hence $\{f_n'(x(:))\}$ is a convergent sequence in $L^{p'}$ for any p'.

LEMMA 6.12. *If* $T(\phi)$ *is linear from* L^p, $(p < \infty)$, *to a subspace* $\mathfrak{M}$ *of finite dimension in* $\mathfrak{X}$, *then there exists an* $x(:)$ *in* $\mathfrak{L}_0^{p'}(\mathfrak{X})$ *defining* $T(\phi)$, *that is,*

$$T(\phi) = \int_0^1 x(s)\phi(s)ds$$

for all ϕ *in* L^p; *and* $x(:)$ *is the limit in* $\mathfrak{L}_0^{p'}(\mathfrak{X})$ *of step-function elements.*

Let x_i, f_i, $(i = 1, 2, \cdots, n)$, be a complete biorthogonal system in $\mathfrak{M}$, so that

$$T(\phi) = \sum_1^n g_i(\phi) \cdot x_i,$$

where $g_i(\phi) = f_i(T(\phi))$. The functional f_i being extensible from $\mathfrak{M}$ to $\mathfrak{X}$, the

last equation implies that $g_i(\phi)$ is a linear functional over L^p; since $p<\infty$, there is an element $\bar{\phi}_i$ in $L^{p'}$ generating $g_i(\phi)$. Set

$$x(s) = \sum_1^n \bar{\phi}_i(s)\cdot x_i. \tag{1}$$

Obviously $x(:)$ is in $\mathfrak{L}_0^{p'}(\mathfrak{X})$; and

$$\begin{aligned}\int_0^1 x(s)\phi(s)ds &= \sum_1^n x_i\cdot \int_0^1 \bar{\phi}_i(s)\phi(s)ds \\ &= \sum x_i g_i(\phi) = T(\phi).\end{aligned}$$

The remainder of the lemma results from approximating, in $L^{p'}$, each $\bar{\phi}_i$ in (1) by a real-valued step-function.

THEOREM 6.1. *A necessary and sufficient condition that a linear operation $V(\phi)$ from L^p, $(1<p<\infty)$, to $\mathfrak{X}$ be c.c. is that it be the strong limit of a sequence of operations defined by step-functions.*

If $V(\phi)$ from L^p to $\mathfrak{X}$ is c.c. and $1<p<\infty$, then ([14], p. 197) V is the limit of operations $\{T_n\}$ where each T_n maps L^p into a finite-dimensional subspace of $\mathfrak{X}$. By Lemma 6.12 there exists a step-function $x_n(:)$ whose corresponding operation V_n is in norm within $1/n$ of T_n, so that $\|V-V_n\|\to 0$.

The converse results from Lemma 6.11 and the fact that the limit under the norm of a sequence of c.c. operations is again c.c.

THEOREM 6.2. *If $x(s)$ is measuralbe and integrable, then $U(f)=f(x(:))$ from $\mathfrak{X}$ to L^1 and $V(\phi)=\int_0^1 x(s)\phi(s)ds$ from L^∞ to $\mathfrak{X}$ are c.c.*

This follows from Theorem 4.3, and Lemma 6.11, and the fact that for $x(:)$ in $\mathfrak{L}_0^1(\mathfrak{X})$

$$\|U\| = \|V\| = ((x))_1.$$

7. Representation of integrals by real-valued kernels. We make the following definition:

DEFINITION 7.1. *$L^{p'q}$, $(1\leqq p'\leqq\infty,\ 1\leqq q\leqq\infty)$, is the class of real-valued functions $F(s,t)$ defined on the unit square and having the following properties:*

(a) *$F(s,t)$ is measurable in (s,t).*

(b) *For each s the function $F(s,:)$ is in L^q.*

(c) *The iterated integral*

$$\int_0^1 \left\{\int_0^1 F(s,t)\bar{\psi}(t)dt\right\}\phi(s)ds \tag{1}$$

exists for every $\bar{\psi}$ in $L^{q'}$ and ϕ in L^p. For $p'=\infty$ the restriction

$$K_F \equiv \underset{\|\bar{\psi}\|_{L^{q'}}=1}{\text{l.u.b.}} \left(\operatorname*{ess.\,sup}_{s} \left| \int_0^1 F(s, t)\bar{\psi}(t)dt \right| \right) < \infty$$

is imposed.

LEMMA 7.21. *If $x(s)$ from $[0, 1]$ to L^q, $(1 \leqq q \leqq \infty)$, is weakly measurable, then there exists a function $F(s, t)$ from the unit square to the reals that is measurable in (s, t) and such that $x(s) = F(s, :)$ as elements in L^q for each s in $[0, 1]$. Any two such "measurable representations" of $x(s)$ differ on at most a set of plane measure zero.*

For $q < \infty$ this is an immediate consequence of Corollary 1.11 and a theorem of Dunford ([9], Theorem 3.1). The case $q = \infty$ results from the preceding, since $x(s)$, if weakly measurable to L^∞, is also weakly measurable when considered as defined to L^1.

THEOREM 7.2. *If $x(:)$ is in $\mathfrak{L}^{p'}(L^q)$, then any measurable representation $F(s, t)$ of $x(s)$ is in $L^{p'q}$. If $q < \infty$ and $F(s, t)$ is in $L^{p'q}$, then the abstract function $x(s) \equiv F(s, :)$ is in $\mathfrak{L}^{p'}(L^q) = \mathfrak{L}_0{}^{p'}(L^q)$.*

Since any $\bar{\psi}$ in $L^{q'}$ defines a functional $f_{\bar{\psi}}$ over L^q, $x(:)$ in $\mathfrak{L}^{p'}(L^q)$ implies that

$$f_{\bar{\psi}}(x(s)) = \int_0^1 F(s, t)\bar{\psi}(t)dt$$

is in $L^{p'}$ and hence that (c) of Definition 7.1 is satisfied; since $F(s, t)$ is a measurable representation of $x(s)$, the remaining two conditions are also fulfilled.

Conversely, if $F(s, t)$ is in $L^{p'q}$, then $x(s) \equiv F(s, :)$ is defined from $[0, 1]$ to L^q; and if $q < \infty$, so that every linear functional $f(\psi)$ over L^q is generated by an element $\bar{\psi}$ of $L^{q'}$, it follows that $f(x(s)) = \int_0^1 F(s, t)\bar{\psi}(t)dt$ is in $L^{p'}$; for if the iterated integral in (c) of Definition 7.1 exists for all ϕ in L^p, then the right-hand side of the above equation is in $L^{p'}$ ([1], p. 85). Thus $x(:)$ is in $\mathfrak{L}^{p'}(L^q) = \mathfrak{L}_0{}^{p'}(L^q)$.

THEOREM 7.3. *If $F(s, t)$ is a measurable representation of an element $x(:)$ of $\mathfrak{L}_0{}^{p'}(L^q)$, then a necessary and sufficient condition that $V(\phi) = \int_0^1 x(s)\phi(s)ds$ from L^p to L^q be expressible as*

$$V(\phi) = \int_0^1 F(s, t)\phi(s)ds \tag{2}$$

is that ϕ in L^p and $\bar{\psi}$ in $L^{q'}$ imply that

$$\int_0^1 \bar{\psi}(t) \left\{ \int_0^1 F(s, t)\phi(s)ds \right\} dt = \int_0^1 \phi(s) \left\{ \int_0^1 F(s, t)\bar{\psi}(t)dt \right\} ds. \tag{3}$$

From Theorem 7.2, $F(s, t)$ is in $L^{p'q}$, so that the right-hand integral in (3) exists. The elements $\bar{\psi}$ of $L^{q'}$ define linear functionals $f_{\bar{\psi}}$ over L^q; hence if (2) is true, then

$$f_{\bar{\psi}}\left(\int_0^1 F(s, t)\phi(s)ds\right) = f_{\bar{\psi}}\left(\int_0^1 x(s)\phi(s)ds\right),$$

or

$$\int_0^1 \bar{\psi}(t)\left\{\int_0^1 F(s, t)\phi(s)ds\right\}dt = \int_0^1 \phi(s)\cdot f_{\bar{\psi}}(F(s, :))ds$$

$$= \int_0^1 \phi(s)\cdot\left\{\int_0^1 F(s, t)\bar{\psi}(t)dt\right\}ds.$$

On the other hand suppose that $x(:)$ is in $\mathfrak{L}_0{}^{p'}(L^q)$, so that $x(:)$ defines the operation $V(\phi) = \int_0^1 x(s)\phi(s)ds$ and suppose that $x(s)$ has a measurable representation $F(s, t)$ satisfying (3). By Theorem 3.4 the operation $\overline{V}$ adjoint to V assigns to each f in $\overline{L^q}$ the linear functional over L^p generated by the element $\bar{\phi} = f(x(:))$ of $L^{p'}$.

Let ϕ be any element of L^p, and let $\mathfrak{F}$ be the set of functionals $f_{\bar{\psi}}$ over L^q generated by elements $\bar{\psi}$ of $L^{q'}$. Then for all $f_{\bar{\phi}} = \overline{V}(f_{\bar{\psi}})$ with $f_{\bar{\psi}}$ in $\mathfrak{F}$

$$f_{\bar{\phi}}(\phi) = \int_0^1 \bar{\phi}(s)\phi(s)ds = \int_0^1 f_{\bar{\psi}}(x(s))\phi(s)ds$$

$$= \int_0^1 \phi(s)\left\{\int_0^1 F(s, t)\bar{\psi}(t)dt\right\}ds,$$

or, if (3) is applied,

$$f_{\bar{\phi}}(\phi) = \int_0^1 \bar{\psi}(t)\cdot\left\{\int_0^1 F(s, t)\phi(s)ds\right\}dt,$$

where $\int_0^1 F(s, t)\phi(s)ds$ must be an element of L^q. Then for a fixed ϕ in L^p,

$$f_{\bar{\phi}}(\phi) = f_{\bar{\psi}}\left(\int_0^1 F(s, t)\phi(s)ds\right)$$

whenever $f_{\bar{\psi}}$ is in $\mathfrak{F}$ and $f_{\bar{\phi}} = \overline{V}(f_{\bar{\psi}})$. But by definition of $\overline{V}$ the equation $f_{\bar{\phi}}(\phi) = f_{\bar{\psi}}(V(\phi))$ holds for all ϕ in L^p; hence, for any fixed ϕ,

$$f_{\bar{\psi}}(V(\phi)) = f_{\bar{\psi}}\left(\int_0^1 F(s, t)\phi(s)ds\right)$$

for all $f_{\bar{\psi}}$ in $\mathfrak{F}$. This implies (2), since the set $\mathfrak{F}$ is a total set of functionals over L^q.

COROLLARY 7.31. *If $x(:)$ in $\mathfrak{L}_0^{p'}(L^q)$ has a measurable representation $F(s, t)$ such that $\int_0^1 \phi(s)\{\int_0^1 |F(s, t)|\bar{\psi}(t)dt\}ds$ exists for every ϕ in L^p and $\bar{\psi}$ in $L^{q'}$, then*

$$\int_0^1 x(s)\phi(s)ds = \int_0^1 F(s, t)\phi(s)ds$$

for all ϕ in L^p.

This follows from Theorem 7.3 and a well known theorem of Tonelli ([2], p. 75).

COROLLARY 7.32. *If $F(s, t)$ is in $L^{\infty q}$, $(q < \infty)$, then*

$$\operatorname*{ess.\,sup.}_S \left(\int_0^1 |F(s, t)|^q dt\right)^{1/q} = K_F < \infty,$$

and

$$V(\phi) = \int_0^1 F(s, t)\phi(s)ds$$

is defined and linear from L^1 to L^q.

By Theorem 7.2, $x(s) \equiv F(s, :)$ is in $\mathfrak{L}^\infty(L^q) = \mathfrak{L}_0^\infty(L^q) = S_{\infty q}$; hence (Definition 3.1)

$$\infty > K_F = \operatorname*{ess.\,sup.}_S \|x(s)\| = \operatorname*{ess.\,sup.}_S \left(\int_0^1 |F(s, t)|^q dt\right)^{1/q}.$$

Clearly $x(:)$ satisfies the hypotheses of Corollary 7.31; hence the operation $V(\phi) = \int_0^1 x(s)\phi(s)ds$ can be written as $V(\phi) = \int_0^1 F(s, t)\phi(s)ds$.

THEOREM 7.4. *If $q < \infty$, if $F(s, t)$, defined and measurable from the unit square to the reals, has $F(s, :)$ in L^q for almost all s, and if*

$$U(\bar{\psi}) = \int_0^1 F(s, t)\bar{\psi}(t)dt$$

is defined from $L^{q'}$ to L^1, then U is linear and c.c.

If in addition $V(\phi) = \int_0^1 F(s, t)\phi(s)ds$ is defined from L^∞ to L^q and $\int_0^1 V(\phi)\cdot\psi dt = \int_0^1 U(\bar{\psi})\phi ds$ for every pair $\bar{\psi}$ in $L^{q'}$, ϕ in L^∞, then $V(\phi)$ is also linear and c.c.*

This results from Definition 7.1, Theorem 7.2, Corollary 5.32, Theorem 7.3, and Theorem 6.2.

8. **Concerning differentiation.** In this section S is taken to be euclidean, and $\alpha(E)$ is the Lebesgue measure function.

* See [9], Theorem 6.1, where for a more restricted type of $F(s, t)$ this operation is shown to be defined and c.c. from L^p to L^q, $(1 < p \leqq \infty, 1 \leqq q < \infty)$.

The relation between the Bochner integral and differentiation is that an additive a.c. BV function of figures is a.e. strongly differentiable if and only if it is a Bochner integral; and the derivative, when it exists a.e., is (Bn) integrable to the original function ([18], p. 410, footnote). However ([6], example 1), an $(\bar{\mathfrak{X}})$ integral may be nowhere weakly differentiable (paragraph 1.14); there is therefore no hope for a theorem similar to the above and dealing with the $(\mathfrak{X})$ integral and either strong or weak derivatives. But if another and more general definition of "derivative" is employed, the corresponding theorem can be stated.

A function $X(R)$ defined to $\mathfrak{X}$ from the figures in a rectangle R_0 of euclidean n-space will be said to be *pseudo-differentiable* or to have a *pseudo-derivative* if there exists an $x(s)$ from R_0 to $\mathfrak{X}$ such that for every f in $\bar{\mathfrak{X}}$ the figure function $f(X(R))$ is differentiable a.e. to the value $f(x(s))$. The function $x(s)$ is a *pseudo-derivative of* $X(R)$.

A necessary and sufficient condition that an additive a.c. function $X(E)$ of measurable sets in R_0 be pseudo-differentiable is that the function $X(E)$ be an $(\mathfrak{X})$ integral.

The sufficiency is obvious; an $(\mathfrak{X})$ integral has the integrand as a pseudo-derivative.

If $X(E)$ has a pseudo-derivative $x(s)$, then if $X(E)$ is additive and a.c., so is $f(X(E))$; hence

$$f(X(E)) = \int_E \left(\frac{d}{ds} f(X(E))\right) ds = \int_E f(x(s))ds,$$

so that $x(s)$ is integrable to the value $X(E)$.

If $\mathfrak{X}$ is weakly complete, the preceding theorem is true for additive a.c. functions of figures. For if $X(R)$ has $x(s)$ for a pseudo-derivative, we have $f(X(R)) = \int_R f(x(s))ds$; if Q is any open set, then $Q = \sum I_n$, where the $\{I_n\}$ are disjoint intervals, and thus

$$\sum |f(X(I_n))| = \sum \left| \int_{I_n} f(x(s))ds \right| \leqq \int_Q |f(x(s))|\, ds < \infty.$$

Since $\mathfrak{X}$ is weakly complete, it follows that $\sum X(I_n)$ has property (O) and therefore that $\sum X(I_n)$ is convergent. Designating this sum by $X(Q)$ we have

$$f(X(Q)) = \sum f(X(I_n)) = \int_Q f(x(s))ds.$$

Finally, if E is an arbitrary measurable set, and if $\{Q_n\}$ are open sets containing E with $\alpha(Q_n) \to \alpha(E)$, then

$$f(X(Q_n)) = \int_{Q_n} f(x(s))ds \to \int_E f(x(s))ds;$$

and, again using the weak completeness of $\mathfrak{X}$, we find that $\{X(Q_n)\}$ converges weakly to an element $X(E)$, and $f(X(E)) = \int_E f(x(s))ds$. Hence $x(s)$ is $(\mathfrak{X})$ integrable to the set-function $X(E)$.

The pseudo-derivative suggests a corresponding "descriptive" definition of an integral: a function $x(s)$ is "integrable" if and only if there exists an additive a.c. function $X(E)$ of measurable sets which has $x(s)$ for a pseudo-derivative; the integral of $x(s)$ over E is $X(E)$. From the above theorem this descriptive integral is the $(\mathfrak{X})$ integral.

If $X(E)$ has a.e. a weak derivative $x(s)$, then $x(s)$ is a pseudo-derivative; hence if $X(E)$ is additive a.c. and weakly differentiable a.e., this weak derivative is $(\mathfrak{X})$ integrable to the value $X(E)$; moreover, by Theorem 1.2 the derivative is measurable, so that $X(E)$ is the $(\mathfrak{X})$ integral of a measurable integrable function, its weak derivative. We have been unable to resolve the converse question: does a measurable integrable function have its integral weakly differentiable a.e.? If so, then by Theorem 5.2 this weak derivative coincides a.e. with the original function.

In conclusion we recall that if $\mathfrak{X}$ is separable and $x(s)$ is essentially bounded and weakly measurable, then $x(s)$ is (Bn) integrable, and $\int_E x(s)ds = (\text{Bn})\int_E x(s)ds$ is therefore strongly differentiable a.e. to $x(s)$. In particular if $\overline{\mathfrak{X}}$ is separable and $x(s)$ is (Bk) integrable and essentially bounded, then $(\text{Bk})\int_E x(s)ds = \int_E x(s)ds$ is strongly differentiable a.e. to $x(s)$. This gives an affirmative answer to a question raised in [6] (p. 378).

9. **Examples.** We give the following examples:

9.1. Let $\mathfrak{X}$ be the space of real functions $\phi(t)$ bounded on $[0, 1]$, with $\|\phi\| = \text{l.u.b.}_t |\phi(t)|$. Let E be a non-measurable set in $S = [0, 1]$; define $x(s) = \theta$ for s in $S - E$, and set $x(s)$ equal to the characteristic function of the point s for s in E. Then $x(s)$ is Graves integrable (see [3]) over every measurable set E to the value θ; hence $x(s)$ is (Bk) integrable and integrable to the same value. The function $x(s)$ is weakly measurable and integrable, yet $\|x(s)\|$ is not measurable. And the set function $X(E) = \int_E x(s)ds \equiv \theta$ is weakly differentiable everywhere to the value θ and thus has two pseudo-derivatives, $x(s)$ and the function identically θ, that differ on a set of positive measure.

9.2. Let $S = [0, 1]$, let $\alpha(E) = mE$, where m is the Lebesgue measure function, and let $\mathfrak{X} = c$, the separable space of convergent sequences of reals. On $J_n = (1/2^n, 1/2^{n-1}]$ define $x(s) = 2^n x_n$, where x_n is the nth unit vector in c.

The function $x(s)$ is in $\mathfrak{L}^1(c)$, that is, $x(s)$ is (D2) integrable, and if $E_n = E \cdot J_n$, then

$$(D2)\quad \int_E x(s)ds = F_E = \{2^1 mE_1, \cdots, 2^n mE_n, \cdots\},$$

an element of $\dot{c} = l^\infty$, where l^∞ is the space of bounded sequences. Here F_E is not a.c., so that, from Theorem 5.3, $x(:)$ is not in $\mathfrak{L}_0^1(c)$;* hence in Corollary 5.31 the hypothesis of weak completeness cannot be omitted. Nor is the set function F_E c.a., since the series $\sum_n F_{J_n}$ in $\bar{c}$ is not even weakly convergent to an element; to prove this it is sufficient to consider the linear functional over $l^\infty = \bar{c}$ defined by Hildebrandt's generalized Stieltjes integral ([13], p. 870) using Banach's measure function ([1], p. 231).

If $y(s) = x(s)$ on $(0, 1]$, if $y(s) = -x(-s)$ for s in $[-1, 0)$, and if $y(0) = \theta$, then $y(s)$ is integrable over $[-1, 1]$ in the sense of Definition 2.1, yet $y(s)$ is not $(\mathfrak{X})$ integrable; this is in contrast to the (Bk) definition, since (Bk) integrability over S implies (Bk) integrability over every measurable subset of S ([6], Theorem 14).

9.3 Let $\{x_n\}$ be a complete orthonormal sequence in L^2, and define $x(s) = 2^n x_n$ on $J_n = (1/2^n, 1/2^n + 1/2^{2n}]$ and $x(s) = \theta$ elsewhere in $[0, 1]$. Then $x(s)$ is (Bn) integrable, and $x(:)$ is in $\mathfrak{L}_0^2(L^2)$, while $\int_0^1 \|x(s)\|^2 ds = \infty$.

This example shows that in Theorem 6.2 we cannot replace $\mathfrak{L}_0^1(\mathfrak{X})$ by $\mathfrak{L}_0^{p'}(\mathfrak{X})$, $(1 \leqq p' < \infty)$. For in $\mathfrak{L}_0^2(L^2)$ there are no step-function elements within $1/2$ of $x(:)$, so that $V(\phi) = \int_0^1 x(s)\phi(s)ds$ from L^2 to L^2 is not c.c. Suppose that $0 < \epsilon < 1/2$ and that $y(s)$ is a step-function taking the values $y_1, \cdots, y_k$. If f_n is the linear functional over L^2 generated by x_n, then $\sum_{n=1}^\infty |f_n(y_i)|^2 < \infty$ for each i; hence there exists an N such that

$$|f_N(y_i)| < \epsilon, \qquad i = 1, \cdots, k.$$

If s is in J_N, then

$$\begin{aligned} |f_N(x(s) - y(s))| &= |f_N(2^N x_N - y_i)| \\ &= 2^N \cdot \left| 1 - \frac{1}{2^N} f_N(y_i) \right| > 2^N \left(1 - \frac{\epsilon}{2^N}\right); \end{aligned}$$

whence

$$\begin{aligned} ((x - y))_2 &\geqq \left(\int_{J_N} |f_N(x - y)|^2 ds \right)^{1/2} \\ &> \left(\frac{1}{2^{2N}} \cdot 2^{2N} \left(1 - \frac{\epsilon}{2^N}\right)^2 \right)^{1/2} \\ &> 1 - \frac{\epsilon}{2^N} > \epsilon. \end{aligned}$$

* This example is similar to one due to Birkhoff, cited indirectly on page 378 of [6].

Thus although every element of the class $S_{p'q}$ ([9], p. 474), composed of functions $x(s)$ measurable to L^q and having $\|x(s)\|^{p'}$ summable, has its $V(\phi)$ c.c., this is no longer true if $S_{p'q}$ is replaced by $\mathfrak{L}_0^{p'}(L^q)$.

9.4. Let $\{x_{ij}\}$ be a complete orthonormal sequence in L^2 arranged in a doubly infinite array, and consider $y_i(s)=x_{ij}$ on $E_{ij}=[j-1/2^i, j/2^i)$, $(j=1, 2, \cdots, 2^i)$, $y_i(0)=y_i(1)=\theta$. Let $x_n(s)=\sum_{i=1}^{n}y_i(s)$. Since $y_i(:)$ is in $\mathfrak{L}_0^2(L^2)$ and $((y_i))_2 \leqq 1/2^{i/2}$, we have

$$((x_n - x_m))_2 \leqq \sum_{i=m+1}^{n} 1/2^{i/2} \to 0$$

as $m, n \to \infty$, so that $\{x_n\}$ is a Cauchy sequence in $\mathfrak{L}_0^p(L^2)$, $(1 \leqq p \leqq 2)$. If there were an $x(:)$ in $\mathfrak{L}_0^p(L^2)$ such that $((x_n-x))_p \to 0$, then (paragraph 3.22) it follows that $((x_n-x))_1 \to 0$ and l.u.b.$_E$ $\|\int_E x_n - \int_E x\| \to 0$ with $1/n$. But as Birkhoff has observed (this example is example 5 of [6]), there exists no $x(s)$ satisfying this last condition. Hence the space $\mathfrak{L}^p(L^2)=\mathfrak{L}_0^p(L^2)$ is not complete for $1 \leqq p \leqq 2$.

If V_n is the operation from $L^{p'}$ to L^2, $(2 \leqq p' \leqq \infty)$, defined by the step-function x_n, then V_n is c.c., and $V=\lim V_n$ is therefore also c.c. But V is generated by no function integrable either (Bn), (Bk), $(\mathfrak{X})$, (D1), or (D2), since the last four integrals have equivalent definitions when $\mathfrak{X}=L^2$ and each of these definitions is more general than that of Bochner.

This example also shows that for step-functions, $((x))_p$ is not topologically equivalent to $(\int_0^1 \|x(s)\|^p ds)^{1/p}$. Moreover, if $X(E)=\lim_n \int_E x_n(s)ds$, then $X(E)$ is the limit of $X_n(E)=\int_E x_n(s)ds$ uniformly with respect to E, and hence is a c.a. and a.c. function from measurable sets to L^2; yet $X(E)$ has a weak derivative on at most a set of measure zero, and from a remark in §8, has no pseudo-derivative. Finally, none of the integral definitions that have been considered here serve to define the general linear or general completely continuous operation from L^2 to L^2.

Conclusion. The following questions are among those we have failed to answer.

Is the integral of a measurable integrable function weakly differentiable a.e.?

Can the hypothesis of measurability be dropped in Theorem 5.3? In Theorem 6.2?

Is the $(\mathfrak{X})$ integral equivalent to that of Birkhoff?

Are step-functions dense in $\mathfrak{L}_0^1(\mathfrak{X})$ for non-separable $\mathfrak{X}$? If so, then every integrable function defines c.c. operations from $\bar{\mathfrak{X}}$ to L^1 and from L^∞ to $\mathfrak{X}$. If the $(\mathfrak{X})$ integral is equivalent to the (Bk) integral, the answer to this question is yes, by paragraph 3.22 above and Theorem 18 of [6].

References

1. S. Banach, *Théorie des Opérations Linéaires*, Warsaw, 1932.

2. S. Saks, *Théorie de l'Intégrale*, Warsaw, 1933.

3. L. M. Graves, *Riemann integration and Taylor's theorem in general analysis*, these Transactions, vol. 29 (1927), pp. 163–177.

4. S. Bochner, *Integration von Funktionen, deren Werte die Elemente eines Vektorräumes sind*, Fundamenta Mathematicae, vol. 20 (1933), pp. 262–276.

5. N. Dunford, *Integration in general analysis*, these Transactions, vol. 37 (1935), pp. 441–453.

6. G. Birkhoff, *Integration of functions with values in a Banach space*, these Transactions, vol. 28 (1935), pp. 357–378.

7. N. Dunford, *Integration of abstract functions*, Bulletin of the American Mathematical Society, abstract 42-3-90.

8. N. Dunford, *Integration of vector-valued functions*, Bulletin of the American Mathematical Society, abstract 43-1-21.

9. N. Dunford, *Integration and linear operations*, these Transactions, vol. 40 (1936), pp. 474–494.

10. G. Birkhoff, *Moore-Smith convergence in general topology*, Annals of Mathematics, (2), vol. 38 (1937), pp. 39–57.

11. W. Orlicz, *Beiträge zur Theorie der Orthogonalentwicklungen* II, Studia Mathematica, vol. 1 (1929), pp. 241–255.

12. W. Orlicz, *Über unbedingte Konvergenz in Funktionenräumen* (I), Studia Mathematica, vol. 4 (1933), pp. 33–37.

13. T. H. Hildebrandt, *On bounded linear functional operations*, these Transactions, vol. 36 (1934), pp. 868–875.

14. T. H. Hildebrandt, *Linear functional transformations in general spaces*, Bulletin of the American Mathematical Society, vol. 37 (1931), pp. 186–212.

15. I. Gelfand, *Sur un lemme de la théorie des espaces linéaires*, Communications de l'Institut des Sciences Mathématiques et mécaniques de l'Université de Kharkoff et la Société Mathématique de Kharkoff, (4), vol. 13, pp. 35–40.

16. G. Fichtenholz and L. Kantorovitch, *Sur les opérations dans l'espace des fonctions bornées*, Studia Mathematica, vol. 5 (1934), pp. 69–98.

17. O. Nikodym, *Sur une généralisation des intégrales de M. J. Radon*, Fundamenta Mathematicae, vol. 15 (1930), pp. 131–179.

18. J. A. Clarkson, *Uniformly convex spaces*, these Transactions, vol. 40 (1936), pp. 396–414.

The University of Virginia,
Charlottesville, Va.

UNIFORMITY IN LINEAR SPACES*

BY
NELSON DUNFORD

INTRODUCTION

The first chapter of this paper concerns itself with questions of uniform boundedness of sets of points in a Banach space and sets of functionals on a Banach space, as well as with a group of closely related resonance theorems. A well known example coming under this heading is the theorem of Toeplitz [36]† stating that $\sup_m \sum_{n=1}^{\infty} |a_{mn}| < \infty$ providing $\eta_m = \sum_{n=1}^{\infty} a_{mn}\xi_n$ converges whenever ξ_n does. Another is the theorem of Hahn [12] stating that if an arbitrary continuous function has the partial sums of its Fourier expansion, with respect to an orthonormal sequence of bounded functions ω_n, essentially bounded, then the sequence $\int \left| \sum_{\nu=1}^{n} \omega_\nu(x)\omega_\nu(t) \right| dt$ is also essentially bounded. Still another is the theorem stating that if the adjoint of an everywhere defined transformation between Banach spaces is everywhere defined, then the transformation is continuous. This was proved, at least for Hilbert space, by von Neumann [21], Stone [34], Tamarkin ([34], p. iv), and Stone and Tamarkin [35] and is probably not usually thought of as a theorem on uniform boundedness.

Most of the results we have in the first chapter are new, others have been proved only in special cases, and some are well known but the proofs heretofore given have been different. Previous methods for discussing questions of uniform boundedness divide themselves into three groups (i) those associated with the names of Lebesgue [18], Banach [2], Hahn [11], and Hildebrandt [13]; (ii) those characterized by the elegant and direct use of the Baire category theorem as in the works of Banach [1], Saks [31], Saks and Tamarkin [32], and others; and (iii) those employed in a recent theorem of Gelfand [9] the proof of which is closely related to that of the category theorem.

The second chapter of this paper is concerned with cases where a weak limiting process implies a strong one. These cases seem to be rather rare but probably more will come to light in the future. The theorem quoted above, stating that the existence of the adjoint implies the continuity of the function, belongs to the class of questions discussed in Chapter II as well as those discussed in Chapter I, while its analogue (Theorem 42) in the Boolean ring of

* Presented to the Society, December 31, 1936, under the title, *Integration of vector-valued functions*; received by the editors September 2, 1937.

† Numbers in brackets refer to the bibliography at the end of the paper.

Lebesgue measurable subsets of (0, 1), which states that an additive vector valued set function $y(e)$ is absolutely continuous providing $\gamma y(e)$ is absolutely continuous for every linear functional γ, is more characteristic of the phenomenon discussed in the second chapter. This is because the continuity of an additive function on a Boolean ring is not a consequence of its boundedness as is the case with linear operations on a Banach space. The theorem on the ring is closely related to the theorem of Orlicz [23] which asserts that for weakly complete Banach spaces the weak unconditional convergence of a series implies its unconditional convergence in the strong sense. Another result in this category which has appeared recently is the theorem of Pitt [26] which states that a bilinear form $\sum \xi_i a_{ij} \eta_j$, if bounded for

$$\|x\|_p = \left(\sum_{i=1}^{\infty} |\xi_i|^p \right)^{1/p} = \|y\|_q = \left(\sum_{j=1}^{\infty} |\eta_j|^q \right)^{1/q} = 1,$$

where $1/p+1/q<1$, is convergent in the sense of Pringsheim as a double series and uniformly for $\|x\|_p=\|y\|_q=1$. This result is not true for $1/p+1/q \geqq 1$. The theorem may be worded in terms of linear operations on l_q and would state that every continuous linear operator on l_q to $l_{p'}$ is completely continuous if $q>p' \geqq 1$. In this form it is closely related to Theorem 71 which has also been proved by Pettis [24] and states that an operator on L^p to L, $(p>1)$, which has the form

$$\psi(s) = \int_0^1 H(s, t)\phi(t)dt,$$

where $H(s, t)$ is in $L^{p'}$ for each s, is necessarily completely continuous. This shows that the expansion of ψ with respect to any complete orthonormal sequence in L is necessarily convergent (in L) uniformly with respect to $\int_0^1 |\phi(t)|^p dt=1$.

With the exception of Pitt's result in the case where $p'>1$, Chapter II contains results considerably broader than those outlined above. Another example of the above type, which falls, however, more naturally into the first chapter is the result that the Riemann integral $\int_0^1 \phi(t)df(t)$, where $f(t)$ has its values in a Banach space, exists for every continuous function ϕ providing the integral $\int_0^1 \phi(t)d\gamma f(t)$ exists for every continuous ϕ and every linear functional γ. Still another is that (Theorem 76) a function $f(z)$ from a domain in the complex z-plane to a complex Banach space is analytic providing $\gamma f(z)$ is analytic for every complex valued linear functional γ.

In Chapter III the results of the preceding chapters are applied to a theory of Lebesgue integration which is broader than those heretofore given

and which furnishes the natural tool for the solution of a general type of moment problem (Theorem 61). The central idea underlying the integral is that a function $f(p)$ (on a measurable set E to a Banach space Y) which has the property that $\gamma f(p)$ is summable for every γ in a closed linear manifold $\Gamma \subset \overline{Y}$, defines uniquely a point in $\overline{\Gamma}$ according to the equation

$$\bar{\gamma}_E(\gamma) = \int_E \gamma f(p)dp.$$

This notion is intimately related to the recent work of Pettis whose manuscript I had the privilege of seeing shortly before I finished the typing of my own. Pettis' paper gives the reader a very interesting discussion of the case where $\Gamma = \overline{Y}$ and $\bar{\gamma}_e$ is in Y for every measurable subset e of E, and this case is undoubtedly one of the most important to be considered. Another case of interest which we have discussed only slightly is the case where Y itself is a conjugate space $\overline{Z}$, and $\Gamma = Z$. Here one always has $\bar{\gamma}_e$ in Y. The integral in this case has been defined by Gelfand [9] but has not, as far as we know, been discussed in any detail.

In Chapter IV a few instances of the general theorems are pointed out.

Notation. The notation we have used is for the most part self explanatory or else explained where it is introduced, but it might help the reader to keep in mind that throughout the paper Y and Z are arbitrary Banach spaces while X is a Banach space subjected to three restrictions imposed at the beginning of Chapter I and later in Chapter II to a fourth restriction. The symbol Γ is always used for a closed linear manifold in $\overline{Y}$, and γ for a point in Γ. Thus in expressions such as $\sup_{\|\gamma\|=1} \gamma y$ it is to be understood, unless explicitly stated to the contrary, that γ is restricted to be in Γ.

We will be dealing with functions $f(t)$ on a class T to a Banach space Y. In connection with these functions the following symbols occur:

$$f(t), \quad f, \quad \gamma f(.), \quad \nu f.$$

The first quite naturally means the value of the function for the argument t. The second is used when we think of f as a point in an abstract space. The third, following the notation of E. H. Moore, we are to interpret as follows: It may be that for a given γ in Γ the numerical function $\gamma f(t)$ on T, when considered as a single entity, is an element of a Banach space X whose elements are numerical functions on T. If so, then $\gamma f(.)$ represents the point in the Banach space X. In the last one ν is a linear functional on a Banach space X whose elements are numerical functions on T. In Chapter III we have shown that the domain of a linear functional ν on X can be extended in a natural way to the class $\mathfrak{X}[Y, \Gamma]$ of all $f(t)$ on T to Y such that $\gamma f(.)$ is in X for every γ

in Γ. The symbol νf is then the value of the function ν for the argument f in $\mathfrak{X}[Y, \Gamma]$. This notation might be confusing in case $\mathfrak{X}[Y, \Gamma]$ were a Banach space (which it sometimes is), and we wished to express the value of a linear functional on $\mathfrak{X}[Y, \Gamma]$ for a particular argument in $\mathfrak{X}[Y, \Gamma]$. But since we shall not have occasion to do this, no confusion should arise. It should be noted that if $X = Y$, then both symbols $\gamma f(.)$ and γf may have a meaning, but the meaning is in general different.

CHAPTER I

1.0. **Uniform boundedness.** Let X be a Banach space composed of numerical functions $\phi(t)$, where t ranges over an abstract set T. Throughout what follows X will be subject to the following conditions:

(1) *If* $\phi_1(t)+\phi_2(t)=\phi(t)$ *for* t *in* T, *then* $\phi_1+\phi_2=\phi$.

(2) *If, for a numerical constant* c, $c\phi_1(t)=\phi(t)$ *for* t *in* T, *then* $c\phi_1=\phi$.

(3) *If* $\phi_n \to \phi$ *and* $\phi_n(t) \to \phi_*(t)$ *for* t *in* T, *then* $\phi=\phi_*$.

Let Y be an arbitrary Banach space and Γ a closed linear manifold in $\overline{Y}$, the space conjugate to Y. The linear space $\mathfrak{X}=\mathfrak{X}[Y, \Gamma]$ is, by definition, the space of all functions $f=f(t)$ on T to Y such that $\gamma f(.)$ is in X for every γ in Γ.

THEOREM 1. *If* f *is in* $\mathfrak{X}[Y, \Gamma]$, *then* $\gamma f(.)$ *is a continuous linear operation on* Γ *to* X. *In other words, there exists a smallest non-negative number* $\|f\|$ *such that*

$$\|\gamma f(.)\|_X \leqq \|f\| \cdot \|\gamma\|, \qquad \gamma \in \Gamma.$$

By (1) and (2) we see that the operation $U(\gamma)=\gamma f(.)$ on Γ to X is additive, that is,

$$U(c_1\gamma_1 + c_2\gamma_2) = c_1U(\gamma_1) + c_2U(\gamma_2),$$

for every pair c_1, c_2 of numerical constants and every pair γ_1, γ_2 of points in Γ. Condition (3) shows that if $\gamma_n \to \gamma^*$ and $U(\gamma_n) \to \phi$, then $U(\gamma^*)=\phi$; thus by a well known theorem ([1], p. 41, Theorem 7), U is a continuous operation from Γ to X.

THEOREM 2. *Let* $f(t)$ *be a function on an arbitrary set* T *to a Banach space* Y. *If*

$$\sup_{t \in T} |\gamma f(t)| < \infty, \qquad \gamma \in \overline{Y},$$

then

$$\sup_{t \in T} \|f(t)\| < \infty.$$

Since $\sup \gamma y = \|y\|$ ([1], p. 55, Theorem 3) where the sup is taken over all γ in $\overline{Y}$ for which $\|\gamma\|=1$, this theorem follows from Theorem 1 by taking $\Gamma=\overline{Y}$ and $X=M^*(T)$ = the space of functions bounded on T.

THEOREM 3. *Let Y be a Banach space and f_t a function on an arbitrary range T to $\overline{Y}$, the space conjugate to Y. If*

$$\sup_{t \in T} | f_t y | < \infty , \qquad y \in Y ,$$

then

$$\sup_{t \in T} \| f_t \| < \infty .$$

This follows from Theorem 1 by taking $X = M^*(T)$ and replacing Y by $\overline{Y}$ and Γ by Y which can be considered as a closed linear manifold in $\overline{\overline{Y}}$ and defined as all $\bar{y}$ in $\overline{\overline{Y}}$ expressible in the form $\bar{y}(\bar{y}) = \bar{y}(y)$.

Occasionally in what follows we shall assume that there is a notion of null set in T. This notion is subject to the single restriction:

(N) *A denumerable sum of null sets is a null set.*

Thus the notion of null set may be defined in terms of a completely additive measure function, in terms of first category sets, or in terms of sets consisting of at most a denumerable number of points, and so on, or may simply mean a void set. Such terms as ess. sup.$_t \phi(t)$ and the Banach space $M(T)$ of essentially bounded functions on T then have a meaning.

THEOREM 4. *Let T be a set (t) of points in which there is a notion of null set satisfying condition* (N). *Let Y be a Banach space for which $\overline{Y}$ is separable. If $f(t)$ on T to Y is such that $\gamma f(t)$ is essentially bounded for every γ in $\overline{Y}$, then $\| f(t) \|$ is essentially bounded.*

In Theorem 1 take $X = M(T)$, $\Gamma = \overline{Y}$; then

$$\text{ess. sup.}_t | \gamma f(t) | \leqq \| f \| \cdot \| \gamma \| .$$

If $\{\gamma_i\}$ is dense in $\overline{Y}$ and T_i is the set in T where

$$| \gamma_i f(t) | > \| f \| \cdot \| \gamma_i \| ,$$

then T_i, and thus $\sum T_i$, is a null set, and

$$| \gamma f(t) | \leqq \| f \| \cdot \| \gamma \|$$

for every γ and every t in $T - \sum T_i$. Thus $\| f(t) \| \leqq \| f \|$ on $T - \sum T_i$.

THEOREM 5. *Let T be a set (t) of points in which there is defined a notion of null set satisfying condition* (N). *Let Y be a separable Banach space. If f_t on T to $\overline{Y}$ is such that $f_t y$ is essentially bounded for every y in Y, then $\| f_t \|$ is essentially bounded.*

The proof is entirely analogous to that of Theorem 4.

THEOREM 6. *Let V be an arbitrary set (v) of points and U a set (u) in which there is a notion of null set satisfying condition* (N). *Let Y be a Banach space for which $\overline{Y}$ is separable. If $f(u, v)$ on UV to Y is such that for each γ in $\overline{Y}$ there is a null set $U_\gamma \subset U$ and a constant M_γ such that*

$$|\gamma f(u, v)| \leqq M_\gamma, \qquad u \in U - U_\gamma, v \in V,$$

then there is a null set $U_0 \subset U$ and a constant M such that

$$\|f(u, v)\| \leqq M, \qquad u \in U - U_0, v \in V.$$

In the product space $T = UV$ of points (u, v), null sets may be defined as sets of the form U_0V, where U_0 is a null set. Theorem 6 is then a corollary of Theorem 4.

THEOREM 7. *Let U and V be as in Theorem 6 and Y a separable Banach space. If $f_{u,v}$ on UV to $\overline{Y}$ is such that for each y in Y there is a null set $U_y \subset U$ and an M_y such that*

$$|f_{u,v}y| \leqq M_y, \qquad u \in U - U_y, v \in V,$$

then there is a null set $U_0 \subset U$ and a constant M such that

$$\|f_{u,v}\| \leqq M, \qquad u \in U - U_0, v \in V.$$

This follows from Theorem 5 as Theorem 6 did from Theorem 4.

The postulate of separability in Theorems 4 and 6 is not entirely necessary. They hold, for example, when $Y = L$, the space of summable functions, and thus $\overline{Y} = M$, the space of essentially bounded and measurable functions, which is not separable. They hold even when $Y = M$ and $\overline{Y}$ is therefore the space of bounded additive set functions ([7], [15]). These facts we shall prove presently.

Indeed they hold (in a more general form in that γ is not arbitrary) for any space Y for which $\overline{Y}$ is the first or second conjugate of a separable space. As the reader will readily see from the argument, the only thing necessary is that $\overline{Y}$ have the following property: Let Y be a Banach space; then the conjugate space $\overline{Y}$ is said to be a *fundamentally separable* space (f.s. space) with determining manifold Γ in case Γ is a separable closed linear manifold in $\overline{Y}$ such that for every y in Y and $\epsilon > 0$ there is a γ in Γ with $\|\gamma\| = 1$ and $\gamma y > \|y\| - \epsilon$; that is, $\sup_{\|\gamma\|=1} \gamma y = \|y\|$ for each y.

THEOREM 8. *If Y is a separable Banach space, then $\overline{Y}$ and $\overline{\overline{Y}}$ are fundamentally separable spaces.*

The space $\overline{Y}$ is an f.s. space, for let y_p be dense in Y, and let γ_p in $\overline{Y}$ be such that ([1], p. 55, Theorem 3)

$$\|\gamma_p\| = 1, \qquad \gamma_p y_p = \|y_p\|.$$

Let Γ be the closed linear manifold determined by γ_p, $(p=1, 2, \cdots)$. Then if $\|y_p - y\| < \epsilon/2$, we have

$$\sup_{\|\gamma\|=1} \gamma y = \sup_{\|\gamma\|=1} [\gamma(y - y_p) + \gamma y_p] > \|y_p\| - \epsilon/2 > \|y\| - \epsilon.$$

In the case of $\overline{Y}$ we can take $\Gamma = Y$.

A determining manifold for M, the space of essentially bounded functions, is the subspace C of continuous functions. This is a consequence of the formula

$$\int f'(t)g(t)dt = \int g(t)df$$

(where $f(t)$ is an absolutely continuous function, $g(t)$ is a continuous function, and the integral on the right is the Riemann-Stieltjes integral) together with the fact that

$$\sup_{|g(t)| \leq 1} \int gdf = \text{(total variation of } f) = \int |f'(t)|\, dt.$$

A determining manifold for the space M of additive set functions $\phi(E)$, with $\|\phi\| =$ total variation of ϕ, is the subspace consisting of the absolutely continuous set functions. This set is isomorphic to L. Similarly, a determining manifold for the space BV of functions f, which are of bounded variation and are normalized so that the total variation of f is the norm of the linear functional $\int gdf$ on C, is the subspace of absolutely continuous functions. As a final example of a non-separable f.s. space we mention the space $\overline{BV}$ [16]. A determining manifold is the space C itself, that is, the set of linear functionals on BV which are expressible in the form

$$\gamma(f) = \int g(t)df,$$

where g is continuous and $\|\gamma\| = \sup_t |g(t)|$.

Theorems 4 and 6 become then, in their more general form, Theorems 9 and 10 below.

THEOREM 9. *Let T be a set (t) of points in which there is a notion of null set satisfying condition* (N). *Let $\overline{Y}$ be an f.s. space with determining manifold Γ. If $f(t)$ on T to Y is such that $\gamma f(t)$ is essentially bounded for every γ in Γ, then $\|f(t)\|$ is essentially bounded.*

THEOREM 10. *Let V be an arbitrary set (v) of points and U a set (u) in which there is a notion of null set satisfying condition* (N). *Let $\overline{Y}$ be an f.s. space with determining manifold Γ. If $f(u, v)$ on UV to Y is such that for each γ in Γ there is a null set $U_\gamma \subset U$ and an M_γ such that*

$$|\gamma f(u, v)| \leqq M_\gamma, \qquad u \in U - U_\gamma, v \in V,$$

then there is a null set $U_0 \subset U$ and a constant M such that

$$\|f(u, v)\| \leqq M, \qquad u \in U - U_0, v \in V.$$

1.1. **Functions of bounded variation.** Let Δ be a finite number of non-overlapping intervals (a_i, b_i) on (a, b). A function $f(P)$ on (a, b) to the Banach space Y is said to be of *bounded variation** on (a, b) in case

$$\sup_{\Delta} \|\Delta f\| < \infty,$$

where $\Delta f = \sum (f(b_i) - f(a_i))$. This is equivalent, in the case of numerical functions, to saying that the sum $\sum |f(b_i) - f(a_i)|$ is bounded in Δ. This notion of bounded variation is too broad for some purposes. For instance in non-separable spaces a function may be of bounded variation and nowhere continuous, for example, $f(t)$ on $(0, 1)$ to the space of bounded functions defined as

$$f(t) = K(s, t) = \begin{cases} 1, & 0 \leqq s \leqq t, \\ 0, & t < s \leqq 1. \end{cases}$$

This cannot happen in separable spaces. But regardless of the space we have the following theorem:

THEOREM 11. *Let f be a function of bounded variation on the interval (a, b) to a Banach space Y. Then the Riemann-Stieltjes integral $\int_a \phi(s) df(s)$ exists for every continuous function ϕ.*

Let $\pi^1 = (\delta_n^1)$, $\pi^2 = (\delta_m^2)$ be two partitions of (a, b) with norm so small that the oscillation of ϕ on any δ_n^1 or δ_m^2 is less than ϵ. Then if $\tau_n^1 \in \delta_n^1$ and $\tau_m^2 \in \delta_m^2$,

$$\sum_n \phi(\tau_n^1)\delta_n^1 f - \sum_m \phi(\tau_m^2)\delta_m^2 f = \sum_n \phi(\tau_n^1) \sum_m (\delta_n^1 \delta_m^2 f) - \sum_m \phi(\tau_m^2) \sum_n (\delta_m^2 \delta_n^1 f)$$
$$= \sum_n \sum_m [\phi(\tau_n^1) - \phi(\tau_m^2)](\delta_n^1 \delta_m^2 f),$$

where $\delta_n^1 \delta_m^2$ is the ordinary product of the two intervals δ_n^1, δ_m^2, and $\delta_n^1 \delta_m^2 f = 0$ in case $\delta_n^1 \delta_m^2$ is empty.

Now to show that this sum is in norm less than or equal to $2\epsilon \sup_\pi \|\sum_\pi \delta f\|$, which is all that is needed, the following lemma will suffice:

* See Gelfand [9] where a corresponding notation is introduced.

LEMMA. *Let $a=(a_1, a_2, \cdots, a_n)$ be a variable point in n-space with norm $\|a\|=\sup_i |a_i|$. Let $y_1, \cdots, y_n$ be n fixed points in a Banach space Y. Then the linear operation*

$$U(a) = \sum_{i=1}^{n} a_i y_i$$

on n-space to Y has its norm

$$\|U\| \leqq 2 \sup_{\sigma} \left\| \sum_{i\epsilon\sigma} y_i \right\|,$$

where σ stands for any set of integers between one and n.

By a well known theorem ([1], p. 55, Theorem 3)

$$\|U\| = \sup_{\|\gamma\|=1} \|\gamma U\|, \quad \text{where} \quad \gamma \epsilon \overline{Y}.$$

Now $\gamma U a = \sum_{i=1}^{n} \gamma a_i y_i$, and

$$\|\gamma U\| = \sum_{i=1}^{n} |\gamma y_i| = \sum_{\sigma_\gamma^+} \gamma y_i - \sum_{\sigma_\gamma^-} \gamma y_i$$

$$\leqq 2\|\gamma\| \sup_{\sigma} \left\| \sum_{i\epsilon\sigma} y_i \right\|,$$

where σ_γ^+ (σ_γ^-) is the set of integers for which $\gamma y_i \geqq 0$ (<0). This completes the proof of the existence of $\int \phi df$.

Consider the space BV_0 of numerical functions $\phi(P)$ of bounded variation on (a, b) with $\phi(a)=0$* and with $\|\phi\| = \int_a |d\phi|$.

THEOREM 12. *If $f(P)$ on (a, b) to the Banach space Y is such that $\gamma f(.)$ is in BV_0 for every γ in $\overline{Y}$, then f is of bounded variation on (a, b).*

In Theorem 1 take $T=(a, b)$, $X=BV_0$ so that

$$|\gamma \Delta f| = |\Delta \gamma f| \leqq \int_a^b |d\gamma f| \leqq \|f\| \cdot \|\gamma\|$$

and thus $\|\Delta f\| \leqq \|f\|$.

We should like to point out another proof of this theorem. In order to put it in the notation of the previous theorems we use t in place of Δ and put $F(t)=tf$. Then $\gamma F(t) = t\gamma f$ is in the space $m(T)$ for every γ in $\overline{Y}$; hence Theorem 12 is a corollary of Theorem 2 applied to $F(t)$.

This latter point of view will save some time for it makes the following three theorems corollaries of Theorems 3, 7, and 10, respectively.

* This restriction is merely for simplicity of notation.

THEOREM 13. *If f_P on (a, b) to $\overline{Y}$ is such that $f_P y$ is in BV_0 for each y in Y, then f_P is of bounded variation on (a, b).*

THEOREM 14. *Let U be a space in which there is a notion of null set satisfying condition* (N), *and let Y be a separable Banach space. If $f_{u,P}$ on $U(a, b)$ to $\overline{Y}$ is such that for each y in Y there is a null set $U_y \subset U$ and an M_y such that*

$$\int_a^b | d_P f_{u,P} y | \leqq M_y, \qquad u \in U - U_y,$$

then there is a null set $U_0 \subset U$ such that $f_{u,P}$ is of bounded variation in P on (a, b) uniformly with respect to u in $U - U_0$.

The v in Theorem 7 plays the role here of t or Δ, and Theorem 7 applied to the function $F_{u,v} = v f_{u,P}$ gives the present theorem. Similarly Theorem 10 gives the theorem:

THEOREM 15. *Let U be a set (u) in which there is a notion of null set satisfying condition* (N). *Let $\overline{Y}$ be an f.s. space with determining manifold Γ. If $f(u, P)$ on $U(a, b)$ to Y is such that for each γ in Γ there is a null set $U_\gamma \subset U$ and an M_γ such that*

$$\int_a^b | d_P \gamma f(u, P) | \leqq M_\gamma, \qquad u \in U - U_\gamma,$$

then there is a null set $U_0 \subset U$ such that $f(u, P)$ is of bounded variation in P on (a, b) uniformly with respect to u in $U - U_0$.

THEOREM 16. *If $f(p)$ on (a, b) to Y is such that the Riemann integral*

$$\int_a^b \phi(p) d\gamma f(p)$$

exists for every continuous function ϕ and every γ in $\overline{Y}$, then the Riemann integral

$$\int_a^b \phi(p) df(p)$$

exists for every continuous function ϕ.

In view of Theorems 11 and 12 it is sufficient to show that the existence of the Riemann integral $\int_a^b \phi(p) d\psi(p)$ of an arbitrary continuous function with respect to the real function ψ implies* $\int_a^b | d\psi | < \infty$.

* This fact is probably well known. It is stated without reference in the introduction of a paper of Pollard [27]. Not knowing where it is proved and because the above proof is another simple application of the principle of uniform boundedness we give the details here.

If ψ is not of bounded variation on (a, b), a sequence π_n of partitions of (a, b) with norm approaching zero can be formed in such a way that each partition π_n is composed of intervals δ_j, δ_j' with

$$\left| \sum_j \delta_j \psi \right| > n.$$

From the principle of uniform boundedness, however, there is a constant M independent of n such that for an arbitrary continuous function $\phi(p)$ we have the inequality

$$\left| \sum_j \phi(\tau_j)\delta_j\psi + \sum_j \phi(\tau_j')\delta_j'\psi \right| \leqq M \sup_{a \leqq p \leqq b} |\phi(p)|,$$

where τ_j (τ_j') is a point in δ_j (δ_j'). For any $n > M$ let ϕ be the continuous function which vanishes on δ_j' and on δ_j has for its graph an isosceles triangle with base δ_j and height one. Then if τ_j is the center of δ_j, we have

$$n < \left| \sum_j \delta_j \psi \right| = \left| \sum_j \phi(\tau_j)\delta_j\psi \right| \leqq M,$$

a contradiction.

In a similar manner it is possible to demonstrate the following theorem:

THEOREM 17. *If f_p on (a, b) to $\overline{Y}$ is such that the Riemann integral*

$$\int_a^b \phi(p)df_p(y)$$

exists for every continuous function ϕ and every y in Y, then the Riemann integral

$$\int_a^b \phi(p)df_p$$

also exists for every continuous function ϕ.

For simplicity of statement we have avoided a generalization which might be made in Theorems 2 to 15 inclusive (except 8 and 11). In all of these theorems the set of $\gamma[y]$ for which $\gamma f(t)[f_t y]$ is bounded, or essentially bounded, need not be assumed to be the whole of $\Gamma[Y]$ but merely a set of second category in $\Gamma[Y]$. We have so far only used Theorem 1 in the case where $X = M(T)$ or $M^*(T)$ and each of these spaces is a special case of a Banach space X where the following condition holds. If $\|\phi_p\| \leqq M$ *for* $p = 1, 2, \cdots$ *and* $\phi_p(t) \to \phi(t)$ *for every* t *in* T, *then* ϕ *is in* X, *and* $\|\phi\| \leqq M$. For Banach spaces X satisfying this condition it is readily shown that if

$\gamma f(.)$ is in X for every γ in a set of the second category in Γ, then $\gamma f(.)$ is in X for every γ in Γ. For the set $\Gamma_0 \subset \Gamma$, where $\gamma f(.)$ is in X, is a linear set and contains a sphere in Γ since it is a second category sum of closed sets $\Gamma_n = \Gamma[\|\gamma f(.)\|_X \leqq n]$. Thus in the case of a space X satisfying the above condition, Theorem 1 can be worded as follows:

THEOREM 1′. *Let X be a Banach space satisfying, besides the conditions* (1), (2), (3), *the condition stated in the preceding paragraph. If the function $f(t)$ on an arbitrary set T to a Banach space Y is such that $\gamma f(.)$ is in X for every γ in a set of second category in the closed linear manifold $\Gamma \subset \overline{Y}$, then f is in $\mathfrak{X}[Y, \Gamma]$ and $\gamma f(.)$ is a continuous operation from Γ to X.*

From this statement of Theorem 1 the generalization mentioned above is obvious.

1.2. **Adjoint operations.** It has been proved by von Neumann [21], Stone ([34], p. 61, Theorem 2.26), Tamarkin ([34], p. iv), and Stone and Tamarkin [35] that an additive operation on Hilbert space to Hilbert space is continuous providing its adjoint is everywhere defined. The corresponding theorem for Banach spaces, together with a number of similar theorems, can be obtained from Theorem 1 by specializing the range T, the function f, and the space X. As the reader will see, these theorems are all of a general type asserting that a limiting process when existing in a weak sense will also exist in a strong sense, and he may therefore expect to find them in Chapter II where such phenomena are discussed. Since the operations involved are additive functions and for such functions continuity is equivalent to boundedness, and since the proofs are entirely characteristic of the preceding proofs, we prefer to group these theorems with those on uniform boundedness.

In this section Y and Z are arbitrary Banach spaces.

By a *determining manifold* in $\overline{Y}$ will be meant a closed linear manifold Γ in $\overline{Y}$ such that

$$\sup_{\|\gamma\|=1} \gamma y = \|y\|, \qquad y \in Y.$$

We shall use the letter Γ for a determining manifold in $\overline{Y}$. It will not be assumed that Γ is separable unless it is so stated. The symbol Γ^* will be used for a set of elements in $\overline{Y}$ (not necessarily in Γ) such that for every γ in Γ there is a sequence γ_n^* of finite linear combinations of elements of Γ^* such that $\gamma_n^* y \to \gamma y$ for every y in Y. The symbols γ, γ^*, μ, with or without subscripts or superscripts, will always stand for points in Γ, Γ^*, $\overline{Z}$, respectively.

We shall be considering functions $y = f(z)$ on a set $D(f)$ (domain of f) dense in Z and with values in Y. The *adjoint* $\bar{f}$ of f is a function on a subset of $\overline{Y}$

with values in $\overline{Z}$. The domain of definition $D(\bar{f})$ of $\bar{f}$ consists of those $\bar{y}$ in $\overline{Y}$ for which there exists a μ such that

$$\bar{y}f(z) = \mu z, \qquad z \in D(f).$$

Since $D(f)$ is dense in Z, μ is unique. The function $\bar{f}$ is then defined by the equation

$$\bar{f}(\bar{y}) = \mu, \qquad \bar{y} \in D(\bar{f}).$$

THEOREM 18. *If the domain $D(f)$ of $y=f(z)$ is dense in Z, and if the adjoint of f is defined for every γ in Γ, then $D(\bar{f})=\overline{Y}$, $\bar{f}$ is continuous, and the domain $D(f)$ may be extended to the whole of Z in such a way that the extended function $f(z)$ is a bounded linear operator with norm the same as that of $\bar{f}$.*

First note that Theorem 2 holds if $\overline{Y}$ is replaced by a determining manifold Γ in $\overline{Y}$. Now in Theorem 2 take T as the set of all z in $D(f)$ with $\|z\| \leqq 1$. By Theorem 2

$$\|f(z)\| \leqq M, \qquad z \in D(f), \|z\| \leqq 1.$$

The rest of the proof is obvious, and we leave it to the reader.

THEOREM 19. *If the adjoint of $y=f(z)$ on Z to Y is defined for every γ^* in Γ^*, then f and $\bar{f}$ are bounded linear operators with the same bound.*

It is a well known corollary of Theorem 3 that if a sequence $\{\mu_n\}$ of linear functionals converges for every z, then the limit is a linear functional. Thus the adjoint $\bar{f}$ is defined for every γ in Γ; and this theorem is a corollary of the preceding one.

THEOREM 20. *If the additive function $y=f(z)$ on Z to Y has the property that $\gamma^*f(z)$ is continuous for every γ^* in Γ^*, then f is continuous.*

This is merely a restatement of Theorem 19, and it is in this form that we prefer to state the further theorems of this type.

We note in passing that for Banach spaces Y which are equivalent to their own conjugates it is possible to define the notion of a *symmetric transformation* and obtain a corollary to Theorem 19 which states that *if an additive symmetric transformation is defined everywhere, then it is continuous.* For example a function $f(y)$ on $D(f) \subset Y$ to Y (where $Y=\overline{Y}$) might be defined as symmetric if it obeys the law

$$y'f(y) = yf(y'), \qquad y, y' \in D(f).$$

THEOREM 21. *If the additive function f_z on Z to $\overline{Y}$ has the property that $f_z y$ is continuous in z for each y in a fundamental set in Y, then f_z is continuous.*

This follows from Theorem 20 by replacing Y by $\overline{Y}$, Γ by Y, and Γ^* by a fundamental set in Y.

THEOREM 22. *If f and f_n, $(n=1, 2, \cdots)$, are additive functions on Z to Y such that $\gamma^* f_n(z)$ is continuous for each γ^* in Γ^* and every integer n, and if*

$$\gamma^* f_n(z) \to \gamma^* f(z), \qquad \gamma^* \in \Gamma^*, z \in Z,$$

then f is continuous.

This follows from Theorems 3 and 20. We leave the details of this theorem, as well as the next, to the reader.

THEOREM 23. *If f_z and f_z^n, $(n=1, 2, \cdots)$, are additive functions on Z to $\overline{Y}$ such that $f_z^n(y)$ is continuous in z for every y in a fundamental set in Y and every integer n, and if*

$$\lim_n f_z^n(y) = f_z(y),$$

for z in Z and y in a fundamental set in Y, then f_z is continuous.

THEOREM 24. *Let S be an arbitrary set (s) of elements, and let $f(z, s)$ on ZS to Y be such that*

(i) *for each γ in Γ and z in Z, $\gamma f(z, s)$ is bounded on S;*

(ii) *for each γ in Γ and s in S, $\gamma f(z, s)$ is a continuous linear functional in z.*

Then $f(z, s)$ is continuous and linear in z uniformly with respect to s; that is, there is a constant M such that

$$\sup_s \|f(z, s)\| \leqq M\|z\|, \qquad z \in Z.$$

Let $T=ZS$, and take X as the space of all real functions $\mu(z, s)$ on T which satisfy the following conditions:

(a) For each z, $\mu(z, s)$ is bounded on S.

(b) For each s, $\mu(z, s)$ is a continuous linear functional on Z.

Then by Theorem 3

$$\sup_s \|\mu(\cdot, s)\| = \sup_s \sup_{\|z\|=1} |\mu(z, s)| < \infty.$$

This constant is taken as the norm of a point in X. Theorem 24 is then a corollary of Theorem 1.

THEOREM 25. *Let S be an arbitrary set of elements (s) and $f_{z,s}$ on ZS to $\overline{Y}$ be such that*

(i) *for y in Y and z in Z, $f_{z,s}(y)$ is bounded on S;*

(ii) *for y in Y and s in S, $f_{z,s}(y)$ is continuous and linear in z.*

Then $f_{z,s}$ is linear and continuous in z uniformly with respect to s, that is, there is a constant M such that

$$\sup_s \|f_{z,s}\| \leqq M\|z\|, \qquad z \in Z.$$

This follows from Theorem 24 by replacing Y by $\overline{Y}$ and Γ by Y.

THEOREM 26. *Let S be an arbitrary set (s) in which there is defined a notion of null set satisfying condition* (N). *Let Z be separable and $f(z, s)$ on ZS to Y such that*

(i) *for each γ in Γ and z in Z there is a constant $M(\gamma, z)$ and a null set $S(\gamma, z)$ such that $|\gamma f(z, s)| \leqq M(\gamma, z)$ on $S - S(\gamma, z)$;*

(ii) *for each γ in Γ and s in S, $\gamma f(z, s)$ is continuous and linear in z.*

Then there is a constant M such that

$$\text{ess. sup.}_s \sup_{\|z\|=1} |\gamma f(z, s)| \leqq M\|\gamma\|;$$

and if Γ is separable, there is a null set S_0 such that

$$\|f(z, s)\| \leqq M\|z\| \qquad s \in S - S_0,\ z \in Z.$$

Take $T = ZS$ and X as the space of all real functions on T of the form $\mu(z, s)$, where

(a) for each s, $\mu(z, s)$ is continuous and linear in z;

(b) for each z, $\mu(z, s)$ is essentially bounded in s.

Then by Theorem 5

$$\text{ess. sup.}_s \|\mu(., s)\| = \text{ess. sup.}_s \sup_{\|z\|=1} |\mu(z, s)| < \infty,$$

and this constant is taken as the norm in X. The first conclusion follows from Theorem 1. If $\{\gamma_j\}$ is dense in Γ, and if S_j the set in S where

$$\sup_{\|z\|=1} |\gamma_j f(z, s)| > M\|\gamma_j\|,$$

then the second conclusion follows by taking $S_0 = \sum_{j=1}^{\infty} S_j$.

THEOREM 27. *Let S and Z be as in Theorem* 26, *and $f_{z,s}$ on ZS to $\overline{Y}$ be such that*

(i) *for each y in Y and z in Z there is a constant $M(y, z)$ and a null set $S(y, z)$ such that $|f_{z,s}(y)| \leqq M(y, z)$ on $S - S(y, z)$;*

(ii) *for each y in Y and s in S, $f_{z,s}(y)$ is continuous and linear in z.*

Then there is a constant M such that

$$\text{ess. sup.}_s \sup_{\|z\|=1} |f_{z,s}(y)| \leqq M\|y\|.$$

If Y is separable, then there is a null set S_0 such that

$$\|f_{z,s}\| \leqq M\|z\| \qquad \text{on} \qquad S - S_0.$$

This follows from Theorem 26 by replacing Y by $\overline{Y}$ and Γ by Y.

The theorems of this section may be conveniently applied to the theory of multilinear forms. One has, for example, the following theorem:

THEOREM 28. *Let Y, Z, Z' be Banach spaces, and $f(z, z')$ on ZZ' to Y be additive in each argument and such that for each γ^* in Γ^*, $\gamma^* f(z, z')$ is continuous in each of the variables z, z' separately. Then there is a constant M such that*

$$\|f(z, z')\| \leqq M\|z\| \cdot \|z'\|.$$

From the continuity of $\gamma^* f(z, z')$ in z and z' separately follows that of $\gamma f(z, z')$ for any γ in Γ. The desired conclusion follows from Theorem 24 by taking S as the unit sphere in Z'. A corollary is the following:

THEOREM 29. *Let $f_{z,z'}$ on ZZ' to $\overline{Y}$ be additive in each argument and such that for each y_0 in a fundamental set in Y, $f_{z,z'}(y_0)$ is continuous in z, z' separately. Then there is a constant M such that*

$$\|f_{z,z'}\| \leqq M\|z\| \cdot \|z'\|.$$

We shall leave further applications of this sort to the reader.

1.3. **Concluding remarks.** Although we shall not have much occasion to use the fact, it might be pointed out that for certain Banach spaces X, the corresponding space $\mathfrak{X}[Y, \Gamma]$ (where Γ is a determining manifold in Y) is itself a Banach space. Suppose the Banach space X satisfies, besides the conditions (1), (2), (3), the further condition:

(A) *If $\phi_n \to \phi$, then $\phi_n(t) \to \phi(t)$ for each t in T.*

It follows from this that $\phi(t) = 0$ on T providing $\phi = 0$, and also that for fixed t, $\nu_t \phi = \phi(t)$ is a linear functional on X. Thus

$$|\phi(t)| \leqq \|\nu_t\| \cdot \|\phi\|, \qquad \phi \in X.$$

This shows that if s is an arbitrary parameter and

$$\lim_{m,n} \|\phi_n^s - \phi_m^s\| = 0$$

uniformly in s, then for each t

$$|\phi_m^s(t) - \phi_n^s(t)| \leqq \|\nu_t\| \cdot \|\phi_m^s - \phi_n\| \to 0$$

uniformly in s. Now suppose Γ is a determining manifold in $\overline{Y}$ and $\{f_m\}$ is a Cauchy sequence of points in $\mathfrak{X}[Y, \Gamma]$. Then

$$\lim_{m,n} \sup_{\|\gamma\|=1} \|\gamma f_m(\cdot) - \gamma f_n(\cdot)\| = 0,$$

and thus

$$\lim_{m,n} \sup_{\|\gamma\|=1} |\gamma f_m(t) - \gamma f_n(t)| = 0, \qquad \text{for each } t,$$

so that the sequence f_m defines uniquely a function $f(t)$ on T to Y such that

$$f_n(t) \to f(t), \qquad t \in T.$$

It is readily shown that f is in $\mathfrak{X}[Y, \Gamma]$ and that $\|f_n - f\| \to 0$ where the norm is taken in the space $\mathfrak{X}$. Thus $\mathfrak{X}[Y, \Gamma]$ is complete. This, together with certain other obvious facts, shows that $\mathfrak{X}[Y, \Gamma]$ is a Banach space.

Now let $U(\gamma)$ be an arbitrary continuous linear operation from Γ to X, and suppose that $Y = \overline{Y}$. Then $\nu_t U(\gamma)$, being continuous in γ, is representable as $\nu_t U(\gamma) = \bar{\gamma}(\gamma)$, where $\bar{\gamma}$ is a point of $\overline{\Gamma}$. By the Hahn-Banach theorem on the extension of linear functionals, there is a point $\bar{y}$ in $\overline{Y}$ such that $\bar{y}(\gamma) = \bar{\gamma}(\gamma)$ for γ in Γ; and since $Y = \overline{Y}$, we have for each t in T a point $f(t)$ in Y such that $\nu_t U(\gamma) = \gamma f(t)$ and thus $U(\gamma) = \gamma f(.)$. This shows that for spaces Y equal to their second conjugates the operation of Theorem 1 is the general linear operator from Γ to X. To summarize we state the following theorem:

THEOREM 30. *Let X be a Banach space which satisfies, besides the conditions* (1), (2), (3), *the further condition* (A). *Let Γ be a determining manifold in $\overline{Y}$. Then the space $\mathfrak{X}[Y, \Gamma]$ with the norm $\|f\|$ of Theorem* 1 *is a Banach space. If in addition $Y = \overline{Y}$, then every continuous linear operator $U(\gamma)$ on Γ to X is expressible in the form*

$$U(\gamma) = \gamma f(.),$$

where f is a point of $\mathfrak{X}[Y, \Gamma]$.

The condition (A) will not be assumed at any other place in this paper.

CHAPTER II

2.0. **Uniform limiting processes.** Most of the applications of Theorem 1 that have been given thus far have been of the general type asserting that boundedness in a weak sense implies boundedness in a strong sense. The applications of the present chapter are to cases where certain limiting processes, when existing in the weak sense, also exist in the strong sense. This is closely related to, and in many cases synonymous with, the statement that the operation of Theorem 1 is not only continuous but completely continuous. This is the case when X is l or L, $\Gamma = \overline{Y}$, and Y is a separable space equal to its second conjugate. Cases other than those in this chapter where a weak limiting process implies a strong one will be found in Chapter IV.

Before proceeding to the results of this chapter we desire to point out the connection between Theorem 32 and a theorem of Orlicz-Banach. Orlicz [23] has shown that for weakly complete spaces the unconditional (that is, absolute in this case) convergence of $\sum_{n=1}^{\infty} \gamma y_n$ for every γ in $\overline{Y}$ implies the un-

conditional convergence of $\sum y_n$. Banach ([1], p. 240) states this theorem in a more general form, namely: If all partial sums of $\sum y_n$ converge weakly to an element in Y, then $\sum y_n$ converges unconditionally. In this form it is not assumed that Y is weakly complete, and the proof can be carried out in a fashion similar to that of Orlicz.*

In view of the known conditions for conditional compactness in l (see, for example, [5], Theorem 2) we can give the following proof of the Orlicz-Banach theorem. Let S be the unit sphere in $\overline{Y}$ and $U(\gamma) = \{\gamma y_n\}$ on $\overline{Y}$ to l which, by Theorem 1, is continuous. It is also completely continuous, for from any sequence in S there is a subsequence γ_i converging on the closed linear manifold determined by y_n, $(n=1, 2, \cdots)$. Since Theorem 1 shows that the sequence $\{U(\gamma_i)\}$ is bounded, to show it convergent weakly and hence in l, it is sufficient to show that $fU(\gamma_i)$ converges for every f in a fundamental set in $\bar{l}$. Such a fundamental set is the set of characteristic functions of sets σ of integers. For such an $f = f_\sigma$,

$$f_\sigma U(\gamma_i) = \sum_{n\in\sigma} \gamma_i y_n = \gamma_i y_\sigma,$$

and this converges since y_σ is in the closed linear manifold determined by y_n. Thus $U(S)$ is conditionally compact; therefore

$$\lim_N \sum_{n=N}^{\infty} |\gamma y_n| = 0$$

uniformly on S, which shows that $\sum_{n\in\sigma} y_n = y_\sigma$ for every set σ of integers.

Theorem 32 to follow is a generalization of the Orlicz-Banach theorem, but the proof even in the case considered by these authors is different from that of Orlicz.

We shall now assume that X is a Banach space of numerical functions $\phi(p)$ on a range P satisfying, besides the conditions (1), (2), and (3) (with T replaced by P), the further condition:

(4) *If ϕ_i is in X, $(i=1, 2, \cdots)$, and $\phi_i(p) \to \phi(p)$ for p in P, and if $\nu\phi_i$ converges for every ν in $\overline{X}$, then ϕ is in X and $\phi_i \to \phi$ in X.*

It is evident that l has this property, for weak and strong convergence in l are equivalent. Another space in which we shall be interested, which satisfies conditions (1) to (4), is the space $L(E) = L(E, \alpha)$. In this symbolism E is an abstract set, and α a completely additive measure function defined on a σ-field $\mathfrak{a}(E)$ of "measurable" subsets of E. The space $L(E)$ is then the space

* Pettis has also given a proof of this theorem; see [24].

of numerical functions on E summable on E with respect to the total variation $|\alpha|$ of α, and the norm of a point ϕ in $L(E)$ is

$$\|\phi\| = \int_E |\phi(p)|\, d|\alpha|.$$

The space $L(E)$ enjoys the property (4), for if $\nu\phi_i$ converges for every ν in $\overline{L(E)}$, we have in particular

$$\lim_i \int_e \phi_i(p)d\alpha$$

existing for every measurable subset e of E. Thus by a theorem of Vitali-Hahn-Saks [31] the integrals are equi-absolutely continuous. It follows immediately that ϕ is in $L(E)$ and $\phi_i \to \phi$ in $L(E)$.

In what follows in this chapter $P=(p)$ and $T=(t)$ are arbitrary sets, X is a Banach space of numerical functions $\phi(p)$ on P which satisfies conditions (1) to (4), and F is a fundamental set in $\overline{X}$. The space is an arbitrary Banach space, Γ is an arbitrary closed linear manifold in $\overline{Y}$, $y(p, t)$ is a function on PT to Y, and Y_0 is the closed linear manifold in Y determined by $y(P, T)$. The symbol $M(X)$ will be used for the space of numerical functions $\phi(p, t)$ in X for each t in T with $\|\phi\| = \sup_t \|\phi(., t)\|_X < \infty$. The following assumptions will sometimes be made:

I. *For every p in P the set $y(p, T)$ is conditionally compact.*

II. *Either Y_0 is separable, or every bounded sequence in Γ contains a subsequence γ_i such that $\gamma_i y_0$ converges for every y_0 in Y_0.*

Note that the first of the two alternative assumptions in II implies the second, and that I implies II in case the range P is denumerable.

In the following theorem the domain $D(U)$ of U is the set of all μ in $\overline{Y}$ for which $U(\mu)=\mu y(p, t)$ is in $M(X)$.

THEOREM 31. *Assume* I *and* II, *and that*

(i) $\Gamma \subset D(U)$, *and*

(ii) *for f in F there is a $y_f(t)$ on T to Y_0 with $y_f(t)$ conditionally compact and*

$$f\gamma y(., t) = \gamma y_f(t), \qquad \gamma \in \Gamma,\ t \in T.$$

It follows that

(iii) *if $\{\gamma_i\}$ is a bounded sequence in Γ, μ a point of $\overline{Y}$, and $\gamma_i y_0 \to \mu y_0$ for y_0 in Y_0, then μ is in $D(U)$ and $U(\gamma_i)$ approaches $U(\mu)$ in $M(X)$;*

(iv) *in case $\Gamma = \overline{Y}$ and $\gamma_i y_0 \to 0$ for y_0 in Y_0, then $U(\gamma_i) \to 0$ in $M(X)$; and*

(v) *$U(\gamma)$ is completely continuous on Γ to $M(X)$.*

To prove (iii) we have by Theorem 1

$$\|U(\gamma_i)\| \leqq \|U\| \sup_i \|\gamma_i\|,$$

so that

$$\sup_i \|\gamma_i y(\cdot, t)\|_X \leqq M.$$

For f in F, $f\gamma_i y(\cdot, t) = \gamma_i y_f(t)$, and this sequence converges for each t. Thus since $\gamma_i y(\cdot, t)$ is a bounded sequence in X, it is weakly convergent. Also $\gamma_i y(p, t) \to \mu y(p, t)$ for every p and t. Hence for each t, $\mu y(\cdot, t)$ is in X and $\gamma_i y(\cdot, t) \to \mu y(\cdot, t)$ in X. Thus

$$\|\mu y(\cdot, t)\|_X = \lim_i \|\gamma_i y(\cdot, t)\|_X \leqq M,$$

so that $U(\mu)$ is in $M(X)$. If $U(\gamma_i)$ does not approach $U(\mu)$ in $M(X)$, there are sequences $i_q \to \infty$ and t_q, and an $\epsilon > 0$ such that

$$0 < \epsilon < \|\lambda_q y(\cdot, t_q)\|_X = \|\phi_q\|_X \leqq 2M,$$

where $\lambda_q = \gamma_{i_q} - \mu$ and $\phi_q = \lambda_q y(\cdot, t_q)$ is a point in X.

To obtain the contradiction we shall show that

(a) $\phi_q(p) \to 0$ for each p in P, and

(b) $f\phi_q$ converges for every f in $\overline{X}$.

Since $V_p \equiv y(p, T)$ is conditionally compact it is totally bounded and is thus covered by a finite number of spheres $K(y_i, \delta)$, $(i = 1, 2, \cdots, n_\delta)$, with centers y_i in v_p and radii δ. There is a q_δ such that

$$|\lambda_q y_i| \leqq \delta, \qquad i = 1, 2, \cdots, n_\delta;\ q \geqq q_\delta,$$

and for each y in V_p there is an i such that $\|y - y_i\|_Y < \delta$. Thus for any y in V_p

$$\begin{aligned} |\lambda_q y| &\leqq |\lambda_q(y - y_i)| + |\lambda_q(y_i)| \\ &\leqq \delta\left[\sup_q \|\lambda_q\| + 1\right], \qquad q \geqq q_\delta. \end{aligned}$$

This shows that $\lambda_q y \to 0$ uniformly on V_p; hence (a) is true. Since $\|\phi_q\|_X \leqq 2M$, it suffices, in proving (b), to show that $f\phi_q$ converges for every f in F. Now from (ii), and the fact that $\gamma_i y(\cdot, t) \to \mu y(\cdot, t)$, we have

$$f\phi_q = f\lambda_q y(\cdot, t_q) = \lambda_q y_f(t_q),$$

and since $y_f(T)$ is conditionally compact, it follows as above that $f\phi_q \to 0$. This completes the proof of (iii).

To prove (iv) note that while $\|\gamma_i\|$ may not be bounded, the sequence $\|\gamma_i\|_0 = \sup_{\|y_0\|=1} |\gamma_i y_0|$ (where y_0 is in Y_0) is bounded. Thus by a theorem of Hahn-Banach ([1], p. 27) on the extension of linear functionals, there is a

sequence $\{\lambda_i\}$ in $\overline{Y}$ such that $\{\lambda_i\}$ is bounded in $\overline{Y}$ and λ_i coincides with γ_i on Y_0. Conclusion (iv) is then a corollary of (iii). In view of II, for every bounded sequence in Γ there is a subsequence γ_i and a functional μ such that $\gamma_i y \to \mu y$ on Y_0. Thus (v) also follows from (iii).

Theorem 31 enables us to state eight theorems some of which will be of use later in the theory of integration. These theorems deal with the following Banach spaces (where $T=(t)$ is an arbitrary set of elements):

E_1 is the space composed of sequences of numerical functions $\phi_n(t)$ such that $\sup_t \sum_{n=1}^{\infty} |\phi_n(t)| < \infty$. The norm is then defined as

$$\|\phi\| = \sup_t \sum_{n=1}^{\infty} |\phi_n(t)|.$$

E_2 is the subset of E_1 for which

$$\lim_{N=\infty} \sup_t \sum_{n=N}^{\infty} |\phi_n(t)| = 0.$$

The norm in E_2 is the same as that in E_1.

E_3: Suppose that in T there is a notion of null set satisfying condition (N). Then E_3 is the set of sequences of numerical functions $\phi_n(t)$ such that ess. $\sup._t \sum_{n=1}^{\infty} |\phi_n(t)| < \infty$. The norm is

$$\|\phi\| = \text{ess. } \sup._t \sum_{n=1}^{\infty} |\phi_n(t)|.$$

E_4 is the subset of E_3 for which

$$\lim_{N=\infty} \text{ess. } \sup._t \sum_{n=N}^{\infty} |\phi_n(t)| = 0.$$

The norm in E_4 is the same as in E_3.

Let E be an abstract set, and α a completely additive numerical measure function defined on a σ-field $\mathfrak{a}(E)$ of "measurable" subsets of E. Let $|\alpha|(e)$ be the total variation of α on e, and suppose that $|\alpha|(E) < \infty$.

E_5 is the space of numerical functions $\phi(p, t)$ on ET such that, for each t in T, $\phi(p, t)$ is summable relative to $|\alpha|$ on E and such that $\sup_t \int_E |\phi(p, t)| d|\alpha| < \infty$. The norm is then

$$\|\phi\| = \sup_t \int_E |\phi(p, t)| \, d|\alpha|.$$

E_6 is the subspace of E_5 for which

$$\lim_{|\alpha|(e)=0} \sup_t \int_e |\phi(p, t)| \, d|\alpha| = 0.$$

The norm in E_6 is the same as for E_5.

E_7: In case there is a notion of null set in T satisfying condition (N), then E_7 is the space of numerical functions on ET such that for almost all t in T, $\phi(p, t)$ is summable with respect to α on E and

$$\|\phi\| = \text{ess. } \sup_t \int_E |\phi(p, t)| \, d|\alpha| < \infty .$$

E_8 is the subspace of E_7 consisting of all functions for which

$$\lim_{|\alpha|(e)=0} \text{ess. } \sup_t \int_E |\phi(p, t)| \, d|\alpha| = 0.$$

The norm in E_8 is the same as in E_7.

In case the set T has but a single element, the first four of these spaces reduce to l while the last four reduce to $L = L(E, \alpha)$.

THEOREM 32. *Let $\{y_n(t)\}$ be a sequence of functions on T to Y such that*

(i) *for every set σ of integers there is a $y_\sigma(t)$ on T to Y_0 such that*

$$\sum_{n\in\sigma} \gamma y_n(t) = \gamma y_\sigma(t), \qquad t \in T, \ \gamma \in \Gamma;$$

(ii) *for every γ in Γ, $U(\gamma) \equiv \{\gamma y_n(t)\}$ is in E_1; and*

(iii) *for every n and σ the sets $y_n(T)$, $y_\sigma(T)$ are conditionally compact.*

Then it follows that

(iv) *the transformation $U(\gamma)$ on Γ to E_1 is completely continuous;*

(v) *for each t in T*

$$\lim_{N=\infty} \sum_{n\geqq N} |\gamma y_n(t)| = 0$$

uniformly with respect to $\|\gamma\| \leqq 1$;

(vi) *the set Γ of linear functionals for which it is assumed that $U(\gamma)$ is defined can be extended to include any γ in $\overline{Y}$ for which there exists a sequence γ_i in Γ with $\|\gamma_i\|$ bounded and $\gamma_i y_n(t) \to \gamma y_n(t)$ for every n and t. If γ_i and γ are such functionals, then*

$$\lim_j \sup_t \sum_{n=1}^{\infty} |\gamma_j y_n(t) - \gamma y_n(t)| = 0;$$

(vii) *in case $\Gamma = \overline{Y}$ and $\gamma_j y \to 0$ for y in Y_0, then*

$$\lim_{j=\infty} \sup_t \sum_{n=1}^{\infty} |\gamma_j y_n(t)| = 0;$$

(viii) *in case Γ is a determining manifold in $\overline{Y}$, then for every set σ of integers the series $\sum_{n\in\sigma} y_n(t)$ converges on T to $y_\sigma(t)$.*

Most of these conclusions follow immediately from the preceding theorem. Here the set P is replaced by the set of integers so that condition II is automatically satisfied. Conclusion (v) follows from (iv) by applying the conditions for compactness in l, and conclusion (viii) follows immediately from (v).

It should be pointed out that if $\Gamma = \overline{Y}$, and Y is weakly complete, then in hypothesis (i) the existence of $y_\sigma(t)$ is implied by the convergence of $\sum_{n\epsilon\sigma}\gamma y_n(t)$. Also if T has but a finite number of points, then (iii) is automatically satisfied. *Thus if* $\overline{Y} = \Gamma$, *if* Y *is weakly complete, and if* $X = l$, *the transformation of Theorem* 1 *is completely continuous.* Similar remarks hold for the following theorems:

THEOREM 33. *Under the hypothesis of the preceding theorem, except that now it is assumed that* $U(\gamma)$ *is in* E_2 *for every* γ *in* Γ, *the conclusions* (v) *and* (viii) *can be strengthened to the following statements*:

(v′) *uniformly with respect to* $\|\gamma\| \leqq 1$,

$$\lim_N \sup_t \sum_N^\infty |\gamma y_n(t)| = 0;$$

(viii′) *in case* Γ *is a determining manifold in* $\overline{Y}$ *then, for every set* σ *of integers, the series* $\sum_{n\epsilon\sigma} y_n(t)$ *converges uniformly on* T *to* $y_\sigma(t)$.

If (v′) were not true, there would exist sequences $N_i \to \infty$, t_i, and γ_i, an $\epsilon > 0$, and a functional γ (perhaps not in Γ) such that

$$\|\gamma_i\| \leqq 1, \quad \gamma_i y_0 \to \gamma y_0, \quad y_0 \epsilon Y_0; \quad 0 < \epsilon < \sum_{N_i}^\infty |\gamma_i y_n(t_i)|, \qquad i = 1, 2, \cdots.$$

By (vi) of the preceding theorem $U(\gamma_i) \to U(\gamma)$ in E_1, and, since E_2 is a closed linear manifold in E_1, $U(\gamma)$ must be in E_2. Thus

$$\begin{aligned}\limsup_t \sum_{N_i}^\infty |\gamma_i y_n(t_i)| &\leqq \lim_t \sup_t \sum_{n=1}^\infty |\gamma_i y_n(t) - \gamma y_n(t)| \\ &= \lim_t \sup_t \sum_{N_i}^\infty |\gamma y_n(t)| = 0,\end{aligned}$$

which is the desired contradiction. Conclusion (viii′) is a corollary of (v′).

THEOREM 34. *Assume* I *and* II *and that* $U(\gamma) = \gamma y(p, t)$ *on* Γ *to* E_5 *is such that*

(i) *for every* e *in* $\mathfrak{a}(E)$ *there is a* $y_e(t)$ *on* T *to* Y_0 *with* $y_e(T)$ *conditionally compact and*

$$\gamma y_e(t) = \int_e \gamma y(p, t) d\alpha, \qquad \gamma \epsilon \Gamma.$$

Then

(ii) *if* $\{\gamma_i\}$ *is a bounded sequence in* Γ, μ *a point of* $\overline{Y}$, *and* $\gamma_i y_0 \to \mu y_0$ *for* y_0 *in* Y_0, *then* $U(\mu)$ *is in* E_5 *and*

$$\lim_i \sup_t \int_E |\gamma_i y(p,t) - \mu y(p,t)|\, d|\alpha| \to 0;$$

(iii) *for each* t

$$\lim_{|\alpha|(e)=0} \int_e |\gamma y(p,t)|\, d|\alpha| = 0$$

uniformly with respect to $\|\gamma\| \leqq 1$;

(iv) U *is completely continuous*;

(v) *in case* $E=(0,1)$ *and* α *is Lebesgue measure, then for each* t

$$\lim_{h=0} \int_0^{1-h} |\gamma y(p+h,t) - \gamma y(p,t)|\, dp = 0$$

uniformly for $\|\gamma\| \leqq 1$.

This follows from Theorem 31. Here F is the set of characteristic functions of measurable sets e. In view of II, conclusion (iii) (which however is not as strong as (iv)) follows from (ii), and (v) is a known condition for compactness in L [37].

THEOREM 35. *If, in addition to the assumptions of the preceding theorem, we assume that for each* γ *in* Γ

$$\lim_{|\alpha|(e)=0} \sup_t \int_e |\gamma y(p,t)|\, d|\alpha| = 0,$$

then this limit exists uniformly with respect to $\|\gamma\| \leqq 1$.

If the conclusion were not true, there would exist sequences t_i, γ_i, e_i (with $|\alpha|(e_i) \to 0$), an $\epsilon > 0$, and a γ (perhaps not in Γ) such that

$$\|\gamma_i\| \leqq 1,$$

$$\gamma_i y_0 \to \gamma y_0, \qquad y_0 \in Y_0,$$

and

$$0 < \epsilon < \int_{e_i} |\gamma_i y(p,t_i)|\, d|\alpha|.$$

By the preceding theorem $U(\gamma)$ is in E_5 and $U(\gamma_i) \to U(\gamma)$ in E_5. But since E_6 is a closed linear manifold in E_5 and $U(\gamma_i)$ is in E_6, it follows that $U(\gamma)$ is in E_6 and thus

$$\limsup_t \int_{e_i} |\gamma_i y(p,t_i)|\, d|\alpha| \leqq \lim_t \sup_t \int_E |\gamma y(p,t) - \gamma_i y(p,t)|\, d|\alpha| + \lim_t \sup_t \int_{e_i} |\gamma y(p,t)|\, d|\alpha| = 0,$$

which is a contradiction.

The following two theorems deal with transformations $U(\gamma)=\gamma y_n(t)$ on Γ to E_3 and E_4, respectively. By T_γ will be understood the set in T for which

$$\sum_{n=1}^{\infty} |\gamma y_n(t)| \leqq \|U\| \cdot \|\gamma\|.$$

Then by Theorem 1, $T-T_\gamma$ is a null set.

THEOREM 36. *Let $y_n(t)$ be a sequence of functions on T to Y, Γ a separable closed linear manifold in $\overline{Y}$ such that*

(i) *for every set σ of integers there is a $y_\sigma(t)$ on T to Y_0 such that for every γ in Γ,*

$$\sum_{n\epsilon\sigma} \gamma y_n(t) = \gamma y_\sigma(t), \qquad t \epsilon T_\gamma;$$

(ii) *for every γ in Γ, $U(\gamma)=\gamma y_n(t)$ is in E_3; and*

(iii) *for every n and σ the sets $y_n(T)$, $y_\sigma(T)$ are conditionally compact.*

Then it follows that

(iv) *in case $\{\gamma_i\}$ is a bounded sequence in Γ with $\gamma_i y_0 \to \mu y_0$ on Y_0, $U(\mu)$ is in E_3 and $U(\gamma_i)\to U(\mu)$ in E_3;*

(v) *U is completely continuous; and*

(vi) *in case Γ is a determining manifold in $\overline{Y}$, there is a null set T_0 such that for every σ we have*

$$\sum_{n\epsilon\sigma} y_n(t) = y_\sigma(t), \qquad t \epsilon T - T_0.$$

THEOREM 37. *If in addition to the conditions of the preceding theorem it is assumed that $U(\gamma)$ is in E_4, then*

$$\lim_N \text{ess. sup.}_t \sum_N^{\infty} |\gamma y_n(t)| = 0$$

uniformly for $\|\gamma\| \leqq 1$. For every σ

$$\sum_{n\epsilon\sigma} y_n(t) = y_\sigma(t)$$

uniformly on $T-T_0$.

To prove Theorems 36 and 37 let $T_0=\sum_{i=1}^{\infty}(T-T_{\gamma_i})$, where $\{\gamma_i\}$ is dense

in Γ so that T_0 is a null set. Now let t be fixed in $T-T_0$, and assume $\epsilon>0$ and N an integer. There is then an i such that

$$\sum_{n=1}^{N} | \gamma y_n(t) - \gamma_i y_n(t) | < \epsilon;$$

hence

$$\sum_{n=1}^{N} | \gamma y_n(t) | \leqq \epsilon + \sum_{n=1}^{\infty} | \gamma_i y_n(t) | \leqq \epsilon + \| U \| \cdot \| \gamma_i \|,$$

which shows that

$$\sup_{t \in T-T_0} \sum_{n=1}^{\infty} | \gamma y_n(t) | \leqq \| U \| \cdot \| \gamma \|, \qquad \gamma \in \Gamma.$$

Theorems 36 and 37 now follow from Theorems 32 and 33 applied to the set $T-T_0$.

The following two theorems deal with a transformation $U(\gamma)=\gamma y(p, t)$ on Γ to E_7 and E_8, respectively. Here T_γ is the set in T for which

$$\int_E | \gamma y(p, t) | \, d | \alpha | \leqq \| U \| \cdot \| \gamma \|.$$

By Theorem 1, $T-T_\gamma$ is a null set.

THEOREM 38. *Assume* I *and* II, Γ *separable, and* $U(\gamma)=\gamma y(p, t)$ *on* Γ *to* E_7 *is such that*

(i) *for every* e *in* $\mathfrak{a}(E)$ *there is a* $y_e(t)$ *on* T *to* Y_0 *with* $y_e(T)$ *conditionally compact such that for* γ *in* Γ

$$\gamma y_e(t) = \int_e \gamma y(p, t) d\alpha \quad \text{on} \quad T_\gamma.$$

Then it follows that

(ii) *if* $\{\gamma_i\}$ *is a bounded sequence in* Γ *and* $\gamma_i y \to \mu y$ *for every* y *in* Y_0, *then* $U(\mu)$ *is in* E_7 *and* $U(\gamma_i) \to U(\mu)$ *in* E_7;

(iii) *for almost all* t

$$\lim_{|\alpha|(e) \to 0} \int_e | \gamma y(p, t) | \, d | \alpha | = 0$$

uniformly for $\|\gamma\| \leqq 1$;

(iv) U *is completely continuous*; *and*

(v) *in case* $E=(0, 1)$ *and* α *is Lebesgue measure, then for almost all* t

$$\lim_{h=0} \int_0^{1-h} | \gamma y(p+h, t) - \gamma y(p, t) | \, dp = 0$$

uniformly for $\|\gamma\| \leqq 1$.

Let γ_i be dense in Γ and $T_0=\sum T-T_{\gamma_i}$. Fix γ in Γ and t_0 in $T-T_0$, and let $\gamma_i \to \gamma$. Then

$$\gamma_i y(p, t_0) \to \gamma y(p, t_0)$$

for every p, and

$$\gamma_i y_e(t) = \int_e \gamma_i y(p, t) d\alpha,$$

providing t is in T_{γ_i}; thus, since t_0 is in T_{γ_i},

$$\gamma_i y_e(t_0) = \int_e \gamma_i y(p, t_0) d\alpha.$$

Since $\gamma_i \to \gamma$, the integrals $\int_e \gamma_i y(p, t_0) d\alpha$ are equi-absolutely continuous (see [31]) and

$$\int_E |\gamma_i y(p, t_0) - \gamma y(p, t_0)| \, d|\alpha| \to 0.$$

Thus

$$\int_E |\gamma y(p, t_0)| \, d|\alpha| = \lim_i \int_E |\gamma_i y(p, t_0)| \, d|\alpha| \leqq \|U\| \cdot \|\gamma_i\| \to \|U\| \cdot \|\gamma\|.$$

This shows that

$$\sup_{t \in T-T_0} \int_E |\gamma y(p, t)| \, d|\alpha| \leqq \|U\| \cdot \|\gamma\|;$$

and the theorem thus follows from Theorem 34 applied to the set $T-T_0$.

In like manner the following theorem follows from Theorem 35:

THEOREM 39. *If in addition to the assumptions of the preceding theorem we assume that for each γ in Γ*

$$\lim_{|\alpha|(e)=0} \text{ess. sup.}_t \int_e |\gamma y(p, t)| \, d|\alpha| = 0,$$

then this limit exists uniformly for $\|\gamma\| \leqq 1$.

THEOREM 40. *Let S be any bounded set in $\overline{Y}$ and $y(e)$ an additive function on $\mathfrak{a}(E)$ to Y such that $y(\mathfrak{a}(E))$ is separable. Then if for each γ in S*

$$\lim_{|\alpha|(e)=0} \gamma y(e) = 0,$$

this same limit holds uniformly for γ in S.

For if not, there exist sequences γ_i and e_i with $|\alpha|(e_i)\to 0$, and a positive ϵ with

$$|\gamma_i y(e_i)| > \epsilon, \qquad i = 1, 2, \cdots.$$

Since $y(\mathfrak{a}(E))$ is separable, there is a subsequence γ_i' of γ_i such that $\gamma_i' y(e)$ converges for each e; thus by a theorem of Saks [31] the functions $\gamma_i' y(e)$ are equi-absolutely continuous, which contradicts the above inequality.

THEOREM 41. *Suppose that the space $\mathfrak{a}(E)$ when metrized by the distance function*

$$(e_1, e_2) = |\alpha|(e_1 + e_2 - e_1e_2)$$

is a separable space. Let Γ be a determining manifold in $\overline{Y}$, and let Γ^* be a set of points in $\overline{Y}$ (not necessarily in Γ) such that for every γ in Γ there is a sequence γ_n^* of finite linear combinations of elements in Γ^* with

$$\lim_n \gamma_n^* y = \gamma y, \qquad y \in Y.$$

Let $y(e)$ be an additive function on $\mathfrak{a}(E)$ to Y such that

(i) *for every sequence $\{e_n\}$ of disjoint sets in $\mathfrak{a}(E)$, $y(\sum e_n)$ is in the closed linear manifold determined by $y(e_n)$;*

(ii) $\lim_{|\alpha|(e)=0} \gamma^* y(e) = 0$, $(\gamma^* \in \Gamma^*)$.

Then

$$\lim_{|\alpha|(e)=0} y(e) = 0.$$

Since for arbitrary γ in Γ

$$\gamma y(e) = \lim_n \gamma_n^* y(e),$$

and each $\gamma_n^* y(e)$ is continuous on the complete metric space $\mathfrak{a}(E)$, the function $\gamma y(e)$ is in Baire's first class and thus continuous at a point. Since it is additive, it must be continuous everywhere.

Let e_n be a sequence of disjoint sets in $\mathfrak{a}(E)$. Then (i) implies that for every set σ of integers $y(\sum_{n\in\sigma} e_n)$ is in the closed linear manifold determined by $y(e_n)$, $(n \in \sigma)$. Further

$$\sum_{n\in\sigma} \gamma y(e_n) = \gamma y\left(\sum_{n\in\sigma} e_n\right), \qquad \gamma \in \Gamma.$$

Thus it follows from Theorem 32 that $y(e)$ is completely additive. From this fact it follows immediately that $\lim_n y(e_n) = y(e)$ if $e_n \to e$ monotonically. Now let $\{e_p\}$ be a sequence dense in $\mathfrak{a}(E)$, and let Y_0 be the closed linear manifold

in Y determined by $y(e)$, where e is a sum of a finite number of the sets e_p. Then Y_0 is separable. Now let e be an arbitrary set in $\mathfrak{a}(E)$, and from a subsequence of $\{e_p\}$ converging in the metric of $\mathfrak{a}(E)$ to e, pick a subsequence e_n' such that

$$\left(e, \prod_{m=1}^{\infty} \sum_{n=m}^{\infty} e_n'\right) = 0,$$

that is, such that

$$e = \prod_{m=1}^{\infty} \sum_{n=m}^{\infty} e_n'$$

except for a null set. Now $y(\sum_{n=m}^{m+p} e_n')$ is in Y_0 and

$$\lim_{p} y\left(\sum_{n=m}^{m+p} e_n'\right) = y\left(\sum_{n=m}^{\infty} e_n'\right),$$

so that for each m, $y(\sum_{n=m}^{\infty} e_n')$ is in Y_0. But

$$\lim_{m} \sum_{n=m}^{\infty} e_n' = \prod_{m=1}^{\infty} \sum_{n=m}^{\infty} e_n'$$

monotonically, and thus

$$y(e) = \lim_{m} y\left(\sum_{n=m}^{\infty} e_n'\right) \text{ is in } Y_0.$$

Hence by the preceding theorem

$$\lim_{|\alpha|(e)=0} y(e) = \lim_{|\alpha|(e)=0} \sup_{\|\gamma\|=1} \gamma y(e) = 0.$$

THEOREM 42. *Suppose that the space $\mathfrak{a}(E)$, when metrized by the distance function*

$$(e_1, e_2) = |\alpha|(e_1 + e_2 - e_1e_2),$$

is a separable space. Let $y(e)$ be an additive function on $\mathfrak{a}(E)$ to Y, and let Γ^* be a set in $\overline{Y}$ such that for every γ in $\overline{Y}$ there is a sequence γ_n^* of finite linear combinations of elements in Γ^* with

$$\lim_{n} \gamma_n^* y = \gamma y, \qquad y \in Y.$$

If

$$\lim_{|\alpha|(e)=0} \gamma^* y(e) = 0, \qquad \gamma^* \in \Gamma^*,$$

then

$$\lim_{|\alpha|(e)=0} y(e) = 0.$$

This follows from the preceding theorem applied to $\Gamma = \overline{Y}$ noting that in this case the equation

$$\sum_n \gamma y(e_n) = \gamma y\Big(\sum_n e_n\Big), \qquad \gamma \in \overline{Y},$$

where $\{e_n\}$ is a sequence of disjoint sets, assures us that hypothesis (i) of the preceding theorem is fulfilled.

CHAPTER III

3.0. **Extension of linear functionals and integration.** In this section the domain of a point ν in $\overline{X}$ will be extended from X to $\mathfrak{X}$; that is, we shall assign a meaning to νf, where ν is a linear functional on X and f is a point in $\mathfrak{X} = \mathfrak{X}[Y, \Gamma]$ and in particular investigate the properties of the integral $\int_a f d\alpha$.

THEOREM 43. *If f is in $\mathfrak{X}[Y, \Gamma]$, ν is in $\overline{X}$, and γ is in Γ, then $\nu\gamma f(.)$ is linear in γ and*

$$|\nu\gamma f(.)| \leqq \|\nu\| \cdot \|\gamma\| \cdot \|f\|.$$

This is a corollary of Theorem 1. The equation

$$\bar{\gamma}(\gamma) = \nu\gamma f(.)$$

then defines uniquely a point $\bar{\gamma}$ in $\overline{\Gamma}$, and we shall define νf as this point in $\overline{\Gamma}$. The notation is perhaps faulty in that it does not show the dependence of νf upon the closed linear manifold Γ, and perhaps $\nu_\Gamma f$ would be preferable. This latter notation will be used in any discussion where the manifold Γ is not fixed. It might be noted here that if f is in $\mathfrak{X}[Y, \Gamma']$, and $\Gamma' \supset \Gamma$, then $\nu_{\Gamma'} f$ is an extension of $\nu_\Gamma f$.

Besides the space $\mathfrak{X}[Y, \Gamma]$ it will sometimes be useful to consider the subclass $\mathfrak{X}_0[Y, \Gamma]$ consisting of all f in $\mathfrak{X}[Y, \Gamma]$ having the property that for every ν in $\overline{X}$ there is a y (depending on ν and f) such that

$$\nu\gamma f = \gamma y, \qquad \gamma \in \Gamma.$$

THEOREM 44. *The space $\mathfrak{X}_0[Y, \Gamma]$ is linear, and in case $\Gamma = \overline{Y}$ it is closed in $\mathfrak{X}[Y, \Gamma]$.*

To see that $\mathfrak{X}_0$ is closed in $\mathfrak{X}$ if $\Gamma = \overline{Y}$, suppose that f_n is in $\mathfrak{X}_0$, f is in $\mathfrak{X}$, and $\|f_n - f\| \to 0$, where the norm is taken in the space $\mathfrak{X}$. This means that

$$\sup_{\|\gamma\| \leq 1} \|\gamma f_n - \gamma f\| \to 0;$$

hence

$$\lim_n \sup_{\|\gamma\| \leq 1} \sup_{\|\nu\| \leq 1} |\nu\gamma f_n - \nu\gamma f| = 0,$$

or

$$\sup_{\|\gamma\|\leqq 1}\sup_{\|\nu\|\leqq 1}\left|\gamma y_n(\nu)-\nu\gamma f\right| = \sup_{\|\nu\|\leqq 1}\left\|y_n(\nu)-\bar{\gamma}(\nu)\right\|_{\bar{\Gamma}}\to 0,$$

where $y_n(\nu)$ is regarded as a point in $\bar{\Gamma}=\overline{\Gamma}$. Thus since Y is closed in $\overline{Y}$ there is a $y(\nu)$ such that $y_n(\nu)\to y(\nu)$ uniformly for $\|\nu\|\leqq 1$ and thus

$$\bar{\gamma}(\nu)\gamma = \nu\gamma f = \gamma y(\nu)$$

for every γ in Γ. The proof that $\mathfrak{X}_0$ is linear will be left to the reader.

THEOREM 45. *If $\Gamma=\overline{\Gamma}$, then a necessary and sufficient condition for a point f in $\mathfrak{X}[Y,\Gamma]$ to be in $\mathfrak{X}_0[Y,\Gamma]$ is that for every ν_0 in a fundamental set $F\subset\overline{X}$ there is a y_0 such that*

$$\nu_0\gamma f(.) = \gamma(y_0), \qquad \gamma\in\Gamma.$$

For if $\nu\gamma f(.)=\gamma(y)$ for every ν in a fundamental set, the same identity holds for every ν in a dense set. Thus if ν is an arbitrary point in $\overline{X}$ and $\|\nu_n-\nu\|\to 0$, where

$$\nu_n\gamma f(.) = \gamma y_n, \qquad \gamma\in\Gamma,$$

we have, using Theorem 43,

$$\|y_m-y_n\| = \sup_{\|\gamma\|\leqq 1}\left|(\nu_n-\nu_m)\gamma f(.)\right| \leqq \|\nu_n-\nu_m\|\cdot\|f\|\to 0,$$

so that if $y=\lim_n y_n$, $\nu\gamma f(.)=\gamma y$.

This proves the sufficiency of the condition, and the necessity is obvious.

3.1. **Integration of numerical functions.** Before proceeding to a discussion of integration of an abstract valued function of an abstract variable it is necessary to set down here certain properties of real summable functions of an abstract variable as well as properties of the Lebesgue integral of such functions.

It is well known from the works of Radon [28], Fréchet [8], Nikodym [22], Ridder [29], and others, that a theory of Lebesgue integration can be developed for real functions of an abstract variable. In fact there are numerous equivalent ways of defining the Lebesgue integral of such a function and several of these are discussed in the paper by Ridder. A basis for such a theory is usually a completely additive family $\mathfrak{a}(E)$ of "measurable" subsets of a given measurable set E and a completely additive real function α. What we have to say here will be based upon the postulates and results in the treatise of Saks ([30], pp. 247–263). Since a completely additive set function is expressible as the difference of two completely additive non-negative set functions, one might restrict the discussion (as Saks does) to the case where α is monotone. We prefer not to assume this. Since only functions which are

summable with respect to $|\alpha|$, the total variation of α, are considered, there can be no meaningless symbols such as $\infty - \infty$ arising in the discussion.

The symbol L^q, $L^q(E)$, or $L^q(E, \alpha)$ will be used for the space of numerical measurable functions $\phi(p)$ on E for which $|\phi(p)|^q$ is summable with respect to $|\alpha|$ on E; and L, $L(E)$, or $L(E, \alpha)$ will be used in place of L^1, $L^1(E)$, $L^1(E, \alpha)$, respectively. Upon introducing the norm

$$\|\phi\| = \left(\int_E |\phi(p)|^q d|\alpha| \right)^{1/q},$$

the space L^q becomes a Banach space.* The triangle property of this norm follows from Minkowski's inequality for denumerable sums together with the fact that a function in L^q can be approached in norm by denumerably valued functions. The completeness of the space can be established in a well known fashion. It suffices to show that a Cauchy sequence ϕ_n in L^q determines a function ϕ such that $\phi_n(p) \to \phi(p)$ approximately with respect to α on E, and secondly that the integrals $\int_e |\phi_n(p)|^q d|\alpha|$ are equi-absolutely continuous. From these two results one readily concludes that ϕ is in L^q and $\|\phi_n - \phi\| \to 0$.

The general linear functional γ on L^q, $(q < \infty)$, is expressible in terms of a point ψ in $L^{q'}$ (where $1/q + 1/q' = 1$) by the formula

$$\gamma\phi = \int_E \psi(p)\phi(p) d|\alpha|,$$

and the norm $\|\gamma\|$ is given by

$$\|\gamma\| = \left(\int_E |\psi(p)|^{q'} d|\alpha| \right)^{1/q'}, \qquad \text{if } q' < \infty,$$

$$= \text{ess. sup. } |\psi(p)|, \qquad \text{if } q' = \infty.$$

To see this we note that by a theorem of Nikodym [22]

$$\gamma\phi_e = \int_e \psi(p) d|\alpha|,$$

where ϕ_e is the characteristic function of the measurable set e. Therefore $\gamma\phi = \int_E \psi(p)\phi(p) d|\alpha|$ for all finitely valued functions ϕ. If ϕ_n is a sequence of finitely valued functions with $\|\phi_n - \phi\| \to 0$, then $\psi(p)\phi_n(p) \to \psi(p)\phi(p)$ approximately with respect to α on E, and $\int_e \psi(p)\phi_n(p) d|\alpha|$ converges for every measurable set e. This shows that $\psi\phi$ is summable on E and that $\gamma\phi = \int_E \psi(p)\phi(p) d|\alpha|$. Furthermore, as the reader can readily show, every

* We assume that $1 \leqq q < \infty$. For $q = \infty$ the space L^q is the space $M(E)$ of essentially bounded and measurable functions.

function in L^q is approachable in norm by finitely valued functions, and thus $\gamma\phi = \int_E \psi(p)\phi(p)d|\alpha|$ for every ϕ in L^q.

To see that ψ is in $L^{q'}$ we proceed as follows. Since $\psi\phi$ is in L for every ϕ in L^q, we have $|\psi(p)|^{1/q} \operatorname{sgn} \psi(p)$ in L^q and thus

$$\int_E |\psi(p)|^{1+1/q} d|\alpha| = \gamma(|\psi(p)|^{1/q} \operatorname{sgn} \psi(p)) \leqq \|\gamma\| \left(\int_E |\psi(p)| \, d|\alpha| \right)^{1/q}$$
$$= \|\gamma\| (\gamma(\operatorname{sgn} \psi))^{1/q} \leqq \|\gamma\|^{1+1/q} [|\alpha|(E)]^{1/q},$$

which shows that $|\psi(p)|^{1/q+1/q^2}$ is in L^q. Thus

$$\int_E |\psi(p)|^{1+1/q+1/q^2} d|\alpha| = \gamma(|\psi(p)|^{1/q+1/q^2} \operatorname{sgn} \psi(p))$$
$$\leqq \|\gamma\| \left(\int_E |\psi(p)|^{1+1/q} d|\alpha| \right)^{1/q}$$
$$\leqq \|\gamma\| \cdot \|\gamma\|^{1/q+1/q^2} [|\alpha|(E)]^{1/q^2}$$
$$= \|\gamma\|^{1+1/q+1/q^2} [|\alpha|(E)]^{1/q^2}.$$

In general

$$\int_E |\psi(p)|^{1+1/q+\cdots+1/q^n} d|\alpha| \leqq \|\gamma\|^{1+1/q+\cdots+1/q^n} [|\alpha|(E)]^{1/q^n}.$$

Taking now first the case where $q>1$, we see that the integrand on the left side of the above inequality approaches $|\psi(p)|^{q'}$; hence by Fatou's lemma

$$\int_E |\psi(p)|^{q'} d|\alpha| \leqq \|\gamma\|^{q'}.$$

By Hölder's inequality $\|\gamma\|^{q'} \leqq \int_E |\psi(p)|^{q'} d|\alpha|$, so that the theorem is established for $q>1$. In case $q=1$ let e_m be the set on which $|\psi(p)| \geqq m$; then

$$m^n |\alpha|(e_m) = \int_{e_m} m^n d|\alpha| \leqq \int_E |\psi(p)|^n d|\alpha| \leqq \|\gamma\|^n |\alpha|(E).$$

That is,

$$(m/\|\gamma\|)^n |\alpha|(e_m) \leqq |\alpha|(E),$$

for every positive number m and every integer n. This shows that $|\alpha|(e_m)=0$ if $m>\|\gamma\|$, or in other words,

$$\text{ess. sup. } |\psi(p)| \leqq \|\gamma\|.$$

On the other hand, it is obvious that

$$\text{ess. sup. } |\psi(p)| \geqq \|\gamma\|,$$

which completes the proof of the theorem.

From well known inequalities it follows that for every ψ in $L^{q'}$ and ϕ in L^q the product $\psi\phi$ is in L; and if ψ is fixed, the function $\int_E \psi(p)\phi(p)d|\alpha|$ is a linear functional on L^q. Thus the conjugate space $\bar{L}^q$ is isometrically isomorphic with $L^{q'}$. To summarize the above facts about L^q we state the following theorem:

THEOREM 46. *Assume $1 \leqq q < \infty$. Then the space $L^q = L^q(E) = L^q(E, \alpha)$ of numerical measurable functions ϕ for which $|\phi|^q$ is summable with respect to $|\alpha|$ on E is a Banach space under the norm*

$$\|\phi\| = \left(\int_E |\phi(p)|^q d|\alpha| \right)^{1/q}.$$

Every linear functional γ on L^q is expressible in the form

$$\gamma\phi = \int_E \psi(p)\phi(p)d|\alpha|,^*$$

where ψ is a point of $L^{q'}$, $(q' = q/(q-1))$, and

$$\|\gamma\| = \left(\int_E |\psi(p)|^{q'} d|\alpha| \right)^{1/q'}, \qquad \text{if } q' < \infty,$$

$$= \text{ess. sup. } |\psi(p)|, \qquad \text{if } q' = \infty.$$

Conversely if ψ is a point of $L^{q'}$, then $\psi\phi$ is summable for every ϕ in L^q, and $\int_E \psi(p)\phi(p)d|\alpha|$ is a linear functional on L^q.

In case $q = \infty$ the general linear functional on L^q is given in terms of an integral with respect to an additive set function of bounded variation [7], [15], but we shall not use this in what follows.

3.2. **Integration of abstract functions.** In case $X = L^q(E, \alpha)$ and the functional ν in $\overline{X}$ is taken as $\nu\phi = \int_e \phi d\alpha$, then the meaning assigned (in §3.0) to νf for f in $\mathfrak{L}^q(E)[Y, \Gamma]$ is to be taken as the definition of $\int_e f d\alpha$.

Our chief interest in this chapter will be the linear space $\mathfrak{L}^q(E)[Y, \Gamma]$, its specializations $\mathfrak{L}^q(E)[Y, \overline{Y}]$, $\mathfrak{L}^q(E)_0[Y, \overline{Y}]$, and $\mathfrak{L}^q(E)[\overline{Y}, Y] = \mathfrak{L}^q(E)_0[\overline{Y}, Y]$, and the integral as a linear operator on these spaces. Here and elsewhere unless explicitly stated to the contrary q is an arbitrary real number with $1 \leqq q \leqq \infty$. The space $\mathfrak{L}(E)_0[Y, \overline{Y}] = \mathfrak{L}^1(E)_0[Y, \overline{Y}]$ includes the various classes of functions called summable or integrable by the authors Graves [10], Hildebrandt [14], Bochner [4], Dunford [6], Birkhoff [3], and, in view of Theorem 45, is identical with the class of summable functions discussed by

* It is also expressible in the form $\gamma\phi = \int_E \xi(p)\phi(p)d\alpha$.

Pettis [24]. The space $\mathfrak{L}(E)[\overline{Y}, Y]$ is the space of summable functionals defined recently by Gelfand [9] for the case $E=(0, 1)$.

We shall begin with a discussion of the general space $\mathfrak{L}^q(E)[Y, \Gamma]$ and the integral on this class. The discussion can then be applied to all of the above cases. A number of the properties of the space $\mathfrak{L}^q(E)[Y, \Gamma]$, and of the integral $\int_E f(p)d\alpha$ on this space, are immediate consequences of the definitions and of Theorem 1. We shall state a few of them first.

THEOREM 47. *The space $\mathfrak{L}^q(E)[Y, \Gamma]$ is a normed linear space, the norm being that of Theorem* 1. *The integral $\int_E f(p)d\alpha$ is a linear (that is, additive and continuous) operation on $\mathfrak{L}^q(E)[Y, \Gamma]$ to $\overline{\Gamma}$.*

THEOREM 48. (i) *If f is in $\mathfrak{L}^q(E)[Y, \Gamma]$, then f is in $\mathfrak{L}^q(e)[Y, \Gamma]$ for every e in $\mathfrak{a}(E)$, the family of measurable subsets of E.*

(ii) *The function f on E to Y is in $\mathfrak{L}^q(E)[Y, \Gamma]$ if and only if $\phi \cdot f$ is in $\mathfrak{L}(E)[Y, \Gamma]$ for every ϕ in $L^{q'}(E)$.*

THEOREM 49. *If γ is in Γ, ϕ in $L^{q'}(E)$, and f in $\mathfrak{L}^q(E)[Y, \Gamma]$, then*

$$\left| \int_E \gamma\phi(p)f(p)d\alpha \right| \leqq \|f\| \cdot \|\phi\| \cdot \|\gamma\|.$$

A corollary of this is the following theorem:

THEOREM 50. *If $1 < q \leqq \infty$, the integral*

$$\int_e \phi(p)f(p)d\alpha$$

is a completely additive and absolutely continuous function on $\mathfrak{a}(E)$ for every f in $\mathfrak{L}^q(E)[Y, \Gamma]$ and ϕ in $L^{q'}(E)$.

Here and elsewhere in the paper a set function $y(e)$ is called *absolutely continuous* in case $y(e) \to 0$ with $|\alpha|(e)$, and *completely additive* in case $y(\sum_n e_n) = \sum_n y(e_n)$ for every sequence $\{e_n\}$ of disjoint measurable sets. In the preceding equality the series on the right must be unconditionally convergent since the left side is independent of the order of the sequence $\{e_n\}$.

THEOREM 51. *If Y is separable and $q > 1$, then*

$$\mathfrak{L}^q(E)[Y, \overline{Y}] = \mathfrak{L}^q(E)_0[Y, \overline{Y}].$$

For if $\gamma_n y \to \gamma y$ for every y in Y, then $\|\gamma_n\|$ is a bounded sequence, and, by the preceding theorem, we have for every e in $\mathfrak{a}(E)$

$$\lim_{|\alpha|(e')=0} \int_{ee'} \gamma_n f(p)d\alpha = 0$$

uniformly in n. Thus

$$\int_e \gamma_n f(p)d\alpha \to \int_e \gamma f(p)d\alpha.$$

By a theorem of Banach ([1], p. 131, Theorem 8) there is a y such that

$$\gamma y = \int_e \gamma f(p)d\alpha, \qquad \gamma \in \overline{Y}.$$

The desired conclusion follows from Theorem 45.

Theorem 51 contains a recent theorem of Krein [17]. Krein has proved that if $f(t)$ on $(0, 1)$ to a Banach space Y has the property that $\gamma f(t)$ is continuous for every γ in $\overline{Y}$, then there is a point y in Y such that

$$\gamma y = \int_0^1 \gamma f(t)d\alpha, \qquad \gamma \in \overline{Y},$$

and has used this result in a discussion of a fixed point theorem. In order to see more clearly the connection between Krein's result and Theorem 51, we should state this theorem in a slightly different form. If we do not want all of $\mathfrak{L}^q(E)[Y, \overline{Y}]$ to be in $\mathfrak{L}^q(E)_0[Y, \overline{Y}]$ but merely want a particular function f in $\mathfrak{L}^q(E)[Y, \overline{Y}]$ to be in $\mathfrak{L}^q(E)_0[Y, \overline{Y}]$, it is sufficient to assume that $f(E-$a null set) is separable (and is thus not necessary to assume the whole of Y to be separable), which is equivalent [24] to assuming that f is measurable. Theorem 51 stated in this new form reads as follows:*

THEOREM 52. *Assume $q>1$, and let f on E to Y be measurable and such that γf is in $L^q(E)$ for every γ in $\overline{Y}$. Then f is in $\mathfrak{L}^q(E)_0[Y, \overline{Y}]$, which means that for every ϕ in $L^{q'}(E)$ there is a y_ϕ such that*

$$\gamma y_\phi = \int_E \phi(p)\gamma f(p)d\alpha, \qquad \gamma \in \overline{Y}.$$

To obtain Krein's result from this it suffices to prove that $f[(0, 1)]$ is separable. Let π_n be a partitioning of $(0, 1)$ into intervals δ_n^m, $(m=1, 2, \cdots, p_n)$, such that the norm of the partitioning approaches zero with $1/n$. If τ_n^m is a point of δ_n^m, and $f_n(t)=f(\tau_n^m)$ on δ_n^m, then $\gamma f_n(t)\to \gamma f(t)$ for every γ and t, and thus $f(t)$ must be in the separable closed linear manifold determined by the points $f(\tau_n^m)$.

In the case considered by Krein more might be said, for the boundedness of $\gamma f(t)$ for every γ in $\overline{Y}$ insures us that the norms $\|f(t)\|$ are bounded, and thus the function f is an absolutely integrable and measurable function.

* Pettis [24] has given a slightly different form of this theorem.

THEOREM 53. *If f is in $\mathfrak{L}(E)[Y, \Gamma]$, and the set of values taken by the integral*

$$\int_e f(p)d\alpha, \qquad e \in \mathfrak{a}(E),$$

is a separable set in $\overline{\Gamma}$, then the integral is a completely additive and absolutely continuous function on $\mathfrak{a}(E)$.

Since the unit sphere in Γ is a bounded set in $\overline{\Gamma}$, this theorem is a corollary of Theorem 40.

THEOREM 54. *Suppose $\mathfrak{a}(E)$ separable,* f in $\mathfrak{L}^a(E)_0[Y, \Gamma]$, where Γ is a determining manifold in $\overline{Y}$, and*

(i) *for any sequence $\{e_n\}$ of disjoint sets, $\int_{\sum e_n} f(p)d\alpha$ is in the closed linear manifold determined by $\int_{e_n} f(p)d\alpha$.*

Then the integral $\int_e f(p)d\alpha$ is a completely additive and absolutely continuous function on $\mathfrak{a}(E)$. Further if $\Gamma = \overline{Y}$, the hypothesis (i) *may be omitted.*

This is a corollary of Theorems 41 and 42. It is not as general as these theorems for there are absolutely continuous set functions which are not indefinite integrals [24].

THEOREM 55. *Let E be the real interval (a, b) and α Lebesgue measure, and let f be in $\mathfrak{L}(E)[Y, \Gamma]$ and ϕ in $L(E)$. Then if*

$$\bar{\gamma}(t) = \int^t f(s)ds, \qquad \Phi(t) = \int^t \phi(s)ds,$$

we have

$$\phi\bar{\gamma} \in \mathfrak{L}(E)[\overline{\Gamma}, \Gamma], \qquad \Phi f \in \mathfrak{L}(E)[Y, \Gamma],$$

and

$$\int_a^b \phi(t)\bar{\gamma}(t)dt = \Phi\bar{\gamma}\Big|_a^b - \int_a^b \Phi(t)f(t)dt.$$

Let $\mathfrak{a}(E')$ be another σ-field of measurable subsets of a measurable set E', and let α' be a completely additive set function on $\mathfrak{a}(E')$.

THEOREM 56. *If the function $f(p, p')$ on EE' to Y belongs to the space† $\mathfrak{L}(EE', \alpha\times\alpha')[Y, \Gamma]$, and if for all p in E, except for a set upon which $|\alpha| = 0$, $f(p, p')$ is in $\mathfrak{L}(E', \alpha')[Y, \Gamma]$, then $\int_{E'} f(p, p')d\alpha'$ is in $\mathfrak{L}(E, \alpha)[\overline{\Gamma}, \Gamma]$, and*

---

* For the case $\Gamma = \overline{Y}$ Pettis proves this theorem without the assumption of separability. His method however can be applied to the theorem as stated here without the assumption of separability.

† For the notion of the product measure on the product space and the general Fubini theorem for real functions see Saks [30], or Łomnicki and Ulam [20].

$$\int_{EE'} f(p, p')d(\alpha \times \alpha') = \int_E d\alpha \int_{E'} f(p, p')d\alpha'.$$

THEOREM 57. *Let f_n be in $\mathfrak{L}^q(E)[Y, \Gamma]$, $(n=1, 2, \cdots)$, and suppose $\int_e f_n(p)d\alpha$ is absolutely continuous for each n. Then if*

$$\lim_n \int_e f_n(p)d\alpha$$

exists for every e in $\mathfrak{a}(E)$, the integrals $\int_e f_n(p)d\alpha$ are equi-absolutely continuous.

The familiar argument of Saks [31] based on the Baire category theorem holds in this environment. For upon setting $F_n(e) = \int_e f_n(p)d\alpha$, the space $\mathfrak{a}(E)$ with metric $(e_1, e_2) = |\alpha|(e_1+e_2-e_1e_2)$ is the sum of the closed sets

$$\mathfrak{a}_q \equiv \mathfrak{a}(E)[\|F_m(e) - F_n(e)\| \leqq \epsilon/3,\ m, n \geqq q],$$

and thus by the Baire theorem, there is an integer q_0 and a sphere $S(e_0, r) \subset \mathfrak{a}_{q_0}$. Let $\delta < r$ be such that

$$\|F_{q_0}(e)\| < \epsilon/3, \qquad |\alpha|(e) < \delta.$$

Thus for $|\alpha|(e) < \delta$ the sets $e_1 = e+(e_0-e)$, $e_2 = e_0-e$ are in $S(e_0, r)$, and

$$\|F_m(e) - F_n(e)\| \leqq \|F_m(e_1) - F_n(e_1)\| + \|F_m(e_2) - F_n(e_2)\| \leqq 2\epsilon/3,$$

for $m, n \geqq q_0$; and for $|\alpha|(e) < \delta$ we have

$$\|F_m(e)\| \leqq \epsilon, \qquad m \geqq q_0.$$

THEOREM 58. *Let f_n be in $\mathfrak{L}(E)[Y, \Gamma]$, $(n=1, 2, \cdots)$, let $f_n(p) \to f(p)$ approximately on E, and let $\int_e f_n(p)d\alpha$ be absolutely continuous for each n. Then the following assertions are equivalent:*

(i) *The limit,* $\lim \int_e f_m(p)d\alpha$, *exists on* $\mathfrak{a}(E)$.

(ii) *The function f is in* $\mathfrak{L}(E)[Y, \Gamma]$ *and*

$$\lim_m \int_e f_m(p)d\alpha = \int_e f(p)d\alpha$$

uniformly on $\mathfrak{a}(E)$.

(iii) $$\lim_{|\alpha|(e)=0} \limsup_m \left| \int_e f_m(p)d\alpha \right| = 0.$$

(iv) $$\lim_{|\alpha|(e)=0} \int_e f_m(p)d\alpha = 0$$

uniformly in m.

The proof of the theorem is not a great deal different than for the case of absolutely integrable and measurable functions $f(p)$ (see Dunford [6], pp. 447–448). There are several differences however; for example, in the present case we do not know that the set $e(n, \epsilon)$ consisting of all the points of e where $\|f_n(p)-f(p)\| \geqq \epsilon$ is a measurable set. It might also be pointed out that in the case of absolutely integrable functions it is necessary to assume that the limit function is summable, while in the broader space $\mathfrak{L}(E)[Y, \Gamma]$ this is part of the conclusion. Before proving the theorem we shall recall the meaning of approximate convergence. The sequence $f_n(p)$ is said to approach $f(p)$ approximately on E in case for every n and $\epsilon>0$ there is a measurable set $e'(n, \epsilon) > e(n, \epsilon)$ such that

$$\lim_n |\alpha|(e'(n, \epsilon)) = 0.$$

If $\|\gamma\| = 1$, and $e(\gamma, n, \epsilon)$ is that part of e on which

$$|\gamma f_n(p) - \gamma f(p)| \geqq \epsilon,$$

then $e(\gamma, n, \epsilon)$ is measurable and

$$e(\gamma, n, \epsilon) \subset e(n, \epsilon) \subset e'(n, \epsilon),$$

which shows that $\gamma f_n(p) \to \gamma f(p)$ in measure. The sequence $f_n(p)$ is said to approach $f(p)$ almost uniformly on E if for every $\epsilon>0$ there are sets E_ϵ and E_ϵ' with E_ϵ' measurable, $E_\epsilon' \supset E-E_\epsilon$, $|\alpha|(E_\epsilon') < \epsilon$, and $f_n(p) \to f(p)$ uniformly on E_ϵ. It is known [6] that if $f_n(p) \to f(p)$ approximately on E, then every subsequence of $f_n(p)$ contains a subsequence which approaches $f(p)$ almost uniformly on E. Now to demonstrate the theorem we shall show the following implications: (iii)$\to$(i)$\to$(iv)$\to$(ii)$\to$(iii). Assuming (iii), we have for every $\epsilon>0$ a $\delta>0$ such that for every measurable e with $|\alpha|(e) < \delta$ there is an n_e such that

$$\left\| \int_e f_n(p) d\alpha \right\| < \epsilon, \qquad n \geqq n_e.$$

Let it first be supposed that $f_n(p)$ converges almost uniformly on E. Then

$$\left\| \int_E (f_m(p) - f_n(p)) d\alpha \right\| \leqq \left\| \int_{E-E'(\delta)} (f_m(p) - f_n(p)) d\alpha \right\| + \left\| \int_{E'(\delta)} (f_m(p) - f_n(p)) d\alpha \right\|,$$

where $E'(\delta) \supset E-E(\delta)$, $|\alpha|(E'(\delta)) < \delta$, and $f_n(p)$ converges uniformly on $E(\delta)$ and hence on $E-E'(\delta)$. There is an m_1 such that

$$\left\| \int_{E-E'(\delta)} (f_m(p) - f_n(p))d\alpha \right\| < \epsilon,$$

$$\left\| \int_{E'(\delta)} (f_m(p) - f_n(p))d\alpha \right\| < 2\epsilon, \qquad m, n \geqq m_1.$$

Thus for $m, n \geqq m_1$ we have

$$\left\| \int_E (f_m(p) - f_n(p))d\alpha \right\| < 3\epsilon.$$

The same proof holds for an arbitrary e in $\mathfrak{a}(E)$. The same conclusion follows if f_n only approaches f approximately, since then every subsequence of $\int_e f_n(p)d\alpha$ contains a subsequence approaching $\int_e f(p)d\alpha$. Thus (iii)→(i). The implication (i)→(iv) follows from the preceding theorem. To show that (iv)→(ii), first note that since $\lim_{|\alpha|(e)=0} \int_e \gamma f_m(p)d\alpha = 0$ uniformly in m, and $\gamma f_m \to \gamma f$ approximately on E, the function γf must be in $L(E)$; hence f is in $\mathfrak{L}(E)[Y, \Gamma]$. Furthermore for each γ we have

(a) $$\lim_m \int_e \gamma f_m(p)d\alpha = \int_e \gamma f(p)d\alpha.$$

To complete the proof of (ii) we have to show that (a) holds uniformly for $\|\gamma\| \leqq 1$ and e in $\mathfrak{a}(E)$. Now

$$\left\| \int_e (f_m(p) - f_n(p))d\alpha \right\| \leqq \left\| \int_{e-E'(m,n,\epsilon)} (f_m(p) - f_n(p))d\alpha \right\| + \left\| \int_{eE'(m,n,\epsilon)} (f_m(p) - f_n(p))d\alpha \right\|,$$

where $E'(m, n, \epsilon)$ is a measurable set covering the set $E(m, n, \epsilon)$ upon which $\|f_m(p) - f_n(p)\| > \epsilon$. Also

$$\left\| \int_{e-E'(m,n,\epsilon)} (f_m(p) - f_n(p))d\alpha \right\| = \sup_{\|\gamma\| \leqq 1} \left| \int_{e-E'(m,n,\epsilon)} \gamma[f_m(p) - f_n(p)]d\alpha \right| \leqq \epsilon |\alpha|(E).$$

There is a $\delta > 0$ such that if $|\alpha|(e) < \delta$,

$$\left\| \int_e (f_m(p) - f_n(p))d\alpha \right\| < \epsilon, \qquad m, n = 1, 2, \cdots,$$

and an m_0 such that

$$|\alpha|(E'(m, n, \epsilon)) < \delta, \qquad n, m \geqq m_0.$$

Consequently, since

$$|\alpha|(eE'(m, n, \epsilon)) \leqq |\alpha|(E'(m, n, \epsilon)) < \delta, \qquad n, m \geqq m_0,$$

we have

$$\left\| \int_e f_m(p)d\alpha - \int_e f_n(p)d\alpha \right\| \leqq \epsilon[1 + |\alpha|(E)], \qquad n, m \geqq m_0;$$

and, in view of (a), this proves (ii). From (ii) it follows that $\int_e f(p)d\alpha$ is absolutely continuous. Thus

$$\lim_{|\alpha|(e)=0} \limsup_m \left\| \int_e f_n(p)d\alpha \right\| = \lim_{|\alpha|(e)=0} \left\| \int_e f(p)d\alpha \right\| = 0;$$

hence (ii) implies (iii), which completes the proof of the theorem.

THEOREM 59. *If f is in $\mathfrak{L}^q(E)[Y, \overline{Y}]$, and if for every e in $\mathfrak{a}(E)$ there is a y in Y such that*

$$\gamma y = \int_e \gamma f(p)d\alpha, \qquad \gamma \in \overline{Y},$$

then for every ϕ in $L^{q'}(E)$ there is a y in Y such that

$$\gamma y = \int_E \gamma\phi(p)f(p)d\alpha.$$

This is a corollary of Theorem 45 in the case where $1 \leqq q < \infty$. The case $q = \infty$ is the case where $\gamma f(p)$ is measurable and $|\gamma f(p)|$ is essentially bounded for every γ in $\overline{Y}$. The integral

$$\int_E \phi(p)f(p)d\alpha,$$

which we know always exists in $\overline{\overline{Y}}$, must be in Y, for if ϕ_n is a sequence of finitely valued functions (that is, functions assuming only a finite number of values) with

$$\|\phi_n - \phi\| = \int_E |\phi_n(p) - \phi(p)| \, d|\alpha| \to 0,$$

then $\int_E \phi_n(p)f(p)d\alpha$ is in Y, and by Theorem 49,

$$\left\| \int_E \phi_n(p)f(p)d\alpha - \int_E \phi(p)f(p)d\alpha \right\| \leqq \|\phi_n - \phi\| \cdot \|f\| \to 0,$$

so that $\int_E \phi(p)f(p)d\alpha$ is also in Y.

THEOREM 60. *For f in $\mathfrak{L}(E)_0[Y, \overline{Y}]$ $(\mathfrak{L}(E)[\overline{Y}, Y])$, the operation $U(\gamma) = \gamma f(.)$ on $\overline{Y}$ $[Y]$ to $L(E)$ is weakly continuous. If f is in $\mathfrak{L}(E)_0[Y, \overline{Y}]$, and $f(E-a$ null set$)$ is a separable set in Y (or if f is measurable*$)$, then U is completely continuous.*

For if $\gamma_i y \to \gamma y$ for every y, the integrals

$$\int_e \gamma_i f(p)d\alpha = \gamma_i y_e$$

are equi-absolutely continuous, and since $\gamma_i f(p) \to \gamma f(p)$ for each p, we have

$$\| U(\gamma_i) - U(\gamma) \| = \int_E | \gamma_i f(p) - \gamma f(p) | d | \alpha | \to 0.$$

A similar proof holds for the case where f is in $\mathfrak{L}(E)[\overline{Y}, Y]$. The complete continuity of U is a corollary of Theorem 34.

THEOREM 61. *If f is in $\mathfrak{L}^q(E)_0[Y, \overline{Y}]$ and ϕ is in $L^q(E)$, then a necessary and sufficient condition that there exist a γ in $\overline{Y}$ with*

(i) $$\|\gamma\| \leqq M, \qquad \gamma f(p) = \phi(p),$$

almost everywhere on E, is that

(ii) $$\left| \int_E \eta(p)\phi(p)d\alpha \right| \leqq M \left\| \int_E \eta(p)f(p)d\alpha \right\|$$

for every η in a set dense in $L^{q'}(E)$.

If there does exist a γ satisfying (i), then (ii) is obviously true. To prove the converse† first note that if (ii) holds for every η in a set dense in $L^{q'}$, then by Theorem 49 it holds for every η in $L^{q'}$. Let Y_0 be the linear manifold in Y determined by

$$\int_E \eta(p)f(p)d\alpha, \qquad \eta \in L^{q'}.$$

Now if

$$y = \int_E \eta(p)f(p)d\alpha = \int_E \eta'(p)f(p)d\alpha,$$

we have

---

* That is, the limit almost everywhere of a sequence of finitely valued functions. In view of a result of Pettis [24] and the fact that γf is measurable, the assumption that $f(E-$a null set$)$ is separable is equivalent to the assumption that f is measurable.

† The proof parallels Banach's discussion of the moment problem ([1], p. 55, Theorem 4).

$$\left|\int_E \eta(p)\phi(p)d\alpha - \int_E \eta'(p)\phi(p)d\alpha\right| \leqq M\left\|\int_E [\eta(p) - \eta'(p)]f(p)d\alpha\right\| = 0,$$

so that

$$\gamma_0(y) = \int_E \eta(p)\phi(p)d\alpha$$

is a well defined additive functional on Y_0. It is also continuous since

$$|\gamma_0 y| = \left|\int_E \eta(p)\phi(p)d\alpha\right| \leqq M\left\|\int_E \eta(p)f(p)d\alpha\right\| = M\|y\|.$$

Hence by the Hahn-Banach theorem on the extension of linear functionals there is a linear functional γ defined on the whole of Y with

$$|\gamma y| \leqq M\|y\|, \qquad y \in Y,$$

$$\gamma y = \gamma_0 y = \int_E \eta(p)\phi(p)d\alpha, \qquad y \in Y_0.$$

Thus for every η in $L^{q'}(E)$

$$\int_E \eta(p)\phi(p)d\alpha = \gamma\int_E \eta(p)f(p)d\alpha = \int_E \eta(p)\gamma f(p)d\alpha.$$

Upon writing $\psi(p) = \phi(p) - \gamma f(p)$, we have

$$\int_E \eta(p)\psi(p)d\alpha = 0, \qquad \eta \in L^{q'}.$$

According to a theorem of Hahn-Sierpiński ([30], p. 249) every E' in $\mathfrak{a}(E)$ is representable as the sum of two disjoint sets E_+', E_-' such that α is non-negative on $\mathfrak{a}(E_+')$ and non-positive on $\mathfrak{a}(E_-')$. It follows immediately that $\psi(p) = 0$ almost everywhere.

THEOREM 62. *Let f_n be in $\mathfrak{L}(E)_0[Y, \overline{Y}]$, $(n = 1, 2, \cdots)$, and $f_n(p) \to f(p)$ approximately on E. Then the following assertions are equivalent:*

(i) *The limit, $\lim_m \int_e f_m(p)d\alpha$, exists on $\mathfrak{a}(E)$.*

(ii) *The function f is in $\mathfrak{L}(E)_0[Y, \overline{Y}]$ and*

$$\lim_m \int_e f_m(p)d\alpha = \int_e f(p)d\alpha$$

uniformly on $\mathfrak{a}(E)$.

(iii) $$\lim_{|\alpha|(e)=0} \limsup_m \left\|\int_e f_m(p)d\alpha\right\| = 0.$$

(iv) *Uniformly with respect to n,*

$$\lim_{|\alpha|(e)=0} \int_e f_n(p)d\alpha = 0.$$

In view of Theorems 54 and 58 there is only one thing to prove here; namely, that f is in $\mathfrak{L}(E)_0[Y, \overline{Y}]$. This however follows immediately from Theorem 59 together with the facts that Y is closed in $\overline{Y}$ and

$$\lim_n \int_e f_n(p)d\alpha = \int_e f(p)d\alpha.$$

THEOREM 63. *Let E be the real interval (a, b), α Lebesgue measure, ϕ real and continuous on (a, b), and f in $\mathfrak{L}(E)_0[Y, \overline{Y}]$ or $\mathfrak{L}(E)[\overline{Y}, Y]$. Then if*

$$\beta(u) = \int_a^u f(v)dv,$$

the Riemann-Stieltjes integral $\int_a^b \phi(u)d\beta(u)$ exists and

$$\int_a^b \phi(u)d\beta(u) = \int_a^b \phi(u)f(u)du.$$

That $\int_a^b \phi(u)d\beta(u)$ exists in the Riemann-Stieltjes sense follows from Theorems 11, 12, and 13; and the equality follows from the corresponding Theorem for numerical functions.

THEOREM 64. *Let E, α be as in the preceding theorem, and let f be in $\mathfrak{L}(E)_0[Y, \overline{Y}]$. If Y is separable, then for every real function Φ absolutely continuous on $a \leqq t \leqq b$ we have almost everywhere,*

$$\frac{d}{ds}\left[\Phi(s)\int_a^s f(t)dt - \int_a^s f(t)\Phi(t)dt\right] = \Phi'(s)\int_a^s f(t)dt.$$

This is not an immediate corollary of Theorem 55 as one might think at first sight. By that theorem F is in $\mathfrak{L}(E)[\overline{Y}, \overline{Y}]$, where

$$F(p) = \Phi'(p)\int_a^p f(t)dt$$

and

$$\int_a^s F(p)dp = \Phi(s)\int_a^s f(t)dt - \int_a^s \Phi(s)f(s)ds.$$

Since f is in $\mathfrak{L}(E)_0[Y, \overline{Y}]$, it follows that F is in $\mathfrak{L}(E)[Y, \overline{Y}]$. This is still not

enough however, for an indefinite integral of a function even in $\mathfrak{L}(E)_0[Y, \overline{Y}]$ is not necessarily differentiable. It will be true that

$$\frac{d}{ds}\int_a^s F(p)dp = F(s)$$

almost everywhere, (which is what we want to show) if $F(p)$ is measurable and $\|F(p)\|$ is summable, that is, if F belongs to the class of absolutely integrable and measurable functions (see Bochner [4]). By Theorem 2, $\int_a^p f(t)dt$ is bounded on (a, b), and for each γ in $\overline{Y}$ the function $\gamma\int_a f(t)dt$ is measurable (in fact absolutely continuous) on (a, b). Since Y is separable, it follows from a theorem of Pettis [24] that $\int_a^p f(t)dt$ is measurable. Thus $F(p)$ is the product of a real summable function and a bounded measurable function and hence is measurable and absolutely integrable.

THEOREM 65. *Let E, α, f, Y be as in the preceding theorem.* If f is defined to be constant outside of (a, b) and $f_h(t) = f(t+h)$, then $\lim_{h=0} f_h - f = 0$, that is,*

$$\lim_{h=0}\int_a^b |\gamma f(t+h) - \gamma f(t)|\, dt = 0$$

uniformly for $\|\gamma\| \leqq 1$.

This follows immediately from Theorem 34 by taking the class T as a class consisting of a single element.

CHAPTER IV

4.0. **Instances.** It is not the purpose of the present chapter to apply the preceding results, as we intend to do that later, but merely to call attention to some special instances of a few of the theorems.

As an instance of Theorem 10 we mention the following theorem which is a generalization of a theorem of Hahn-Steinhaus [12], given by Saks and Tamarkin [32].

THEOREM 66. *Let V be an arbitrary set (v) and U a set (u) in which there is a notion of null set satisfying condition* (N). *If for every u in U (or almost all u in U) and every v in V the function $H(t, u, v)$ is of bounded variation on $a \leqq t \leqq b$ and normalized so that*†

$$H(t, u, v) = 1/2[H(t+0, u, v) + H(t-0, u, v)],$$

---

* In both of these theorems it is not entirely necessary that Y be separable. All that is needed is that $f(E-\text{a null set})$ be separable which, in view of Pettis' result, is equivalent to saying that f is measurable.

† Or normalized in any other manner so that its total variation is the same as the norm of the linear functional it defines on C.

then if, for every continuous function ϕ defined on (a, b), there is a null set $U_\phi \subset U$ and an M_ϕ such that

$$\int_a^b \phi(t) d_t H(t, u, v) < M_\phi, \qquad u \in U - U_\phi, v \in V,$$

there will be a null set $U_0 \subset U$ and a constant M such that

$$\int_a^b | d_t H(t, u, v) | < M, \qquad u \in U - U_0, v \in V.$$

This corresponds to taking $Y = BV$, $\Gamma = C$. Another choice might be $Y = M$, $\Gamma = L$. This would yield the theorem:

THEOREM 67. *Let U and V be as in the preceding theorem, and $H(t, u, v)$ essentially bounded in t for almost all u and all v. Then if for every summable function ϕ there is a null set $U_\phi \subset U$ and a constant M_ϕ such that*

$$\int_a^b \phi(t) H(t, u, v) dt < M, \qquad u \in U - U_\phi, v \in V,$$

there is a null set $U_0 \subset U$ and a constant M such that

$$\text{ess. sup.}_t | H(t, u, v) | < M, \qquad u \in U - U_0, v \in V.$$

THEOREM 68. *Let $\mathfrak{a}(E)$, $\mathfrak{a}(E')$ be two families consisting of all measurable subsets of the measurable sets E and E', respectively. Let α, α' be positive, finite, completely additive set functions on $\mathfrak{a}(E)$, $\mathfrak{a}(E')$, respectively. Let $K(p, p')$ be a real function on EE' such that the integral*

$$\int_{e'} K(p, p') d\alpha', \qquad e' \in \mathfrak{a}(E'),$$

exists for almost all (with respect to α) p in E and is essentially bounded and measurable on E. If for every ϕ in $L(E, \alpha)$ there is a constant M_ϕ such that

$$\left| \int_E \phi(p) d\alpha \int_{e'} K(p, p') d\alpha' \right| \leqq M_\phi \alpha'(e'),$$

then for every ψ in $L(E', \alpha')$ the integral

$$\int_{E'} K(p, p') \psi(p') d\alpha'$$

exists for almost all p, is essentially bounded and measurable in p, and

$$\lim_{\psi = 0} \text{ess. sup.}_p \left| \int_{E'} K(p, p') \psi(p') d\alpha' \right| = 0.$$

In Theorem 18 take $Z=L(E', \alpha')$, $Y=L^{\infty}(E, \alpha)$, $\Gamma=L(E, \alpha)$, and

$$f(z) = \int_{E'} K(p, p')z(p')d\alpha'.$$

By assumption $f(z)$ is defined for all finitely valued functions, so that $D(f)$ is dense in $L(E', \alpha')$. Also by assumption

$$| \gamma f(z) | \leqq M_{\gamma}\|z\|$$

providing z is the characteristic function of a measurable set e'. It is readily shown that this same inequality holds for any finitely valued z, and the adjoint of f is thus defined for every γ in Γ. Then by Theorem 18, $f(z)$ can be extended to be linear and bounded on $L(E', \alpha')$. Let z_n be finitely valued functions approaching z in $L(E', \alpha')$. Then there exists a set $E_0 \epsilon \mathfrak{a}(E)$ with $\alpha(E-E_0)=0$ such that

$$\int_{E'} K(p, p')z_n(p')d\alpha'$$

converges uniformly with respect to p in E_0. Furthermore for each p in E_0

$$K(p, p')z_n(p') \to K(p, p')z(p)$$

approximately with respect to α' on E', and

$$\operatorname{ess.\,sup.}_{p} \left| \int_{e'} K(p, p')z_n(p')d\alpha' \right| = \|f(z_n\psi_{e'})\| \leqq \|f\| \int_{e'} | z_n(p') | d\alpha',$$

where $\psi_{e'}$ is the characteristic function of e'. This shows that the integrals $\int_{e'}K(p, p')z_n(p')d\alpha'$ are, at least for p in a set of measure $\alpha(E)$, equi-absolutely continuous, and thus $\int_E K(p, p')z(p')d\alpha'$ exists almost everywhere and is essentially bounded. The last conclusion follows from Theorem 18.

THEOREM 69. *Let Z be a Banach space and $f(z)=K(z, t)$ an additive function on Z to $L_q(0, 1)$. If*

$$\lim_{z=0} \int_0^s K(z, t)dt = 0, \qquad 0 \leqq s \leqq 1,$$

then

$$\lim_{z=0} \operatorname{ess.\,sup.}_{t} | K(z, t) | = 0, \qquad \text{if } q = \infty,$$

$$\lim_{z=0} \left(\int_0^1 | K(z, t) |^q dt \right)^{1/q} = 0, \qquad \text{if } 1 \leqq q < \infty.$$

In case $1<q\leqq\infty$, this is Theorem 21 applied to $L^{q'}(0, 1)$. For $q=1$ one

can use Theorem 20 with $Y=L$, $\Gamma=C=$continuous functions, $\Gamma^*=$characteristic functions of intervals.

THEOREM 70. *Let Z be a Banach space, and $f(z)=K(z, t)$ an additive function on Z to the space $\overline{C}$ (where C is the space of continuous functions on $(0, 1)$). If*

$$\lim_{z=0} K(z, t) = 0, \qquad 0 \leqq t \leqq 1,$$

then

$$\lim_{z=0} \int_0^1 | d_t K(z, t) | = 0.$$

This is Theorem 20 applied to the case where $Y=\overline{C}$, $\Gamma=C$, $\Gamma^*=$characteristic functions of intervals.

Assume $p>1$; then an instance of Theorem 34 (or 60), if we use the fact that if $Y=\overline{Y}$, then $\mathfrak{L}(E)[Y, \overline{Y}]=\mathfrak{L}(E)_0[Y, \overline{Y}]$, is the following theorem:

THEOREM 71. *If for almost all s in $(0, 1)$ the function $H(s, t)$ is in $L^{p'}(0, 1)$, and if*

$$U(\phi) = \psi(s) = \int_0^1 H(s, t)\phi(t)dt$$

is in $L(0, 1)$ for every ϕ in $L^p(0, 1)$, then the operation $U(\phi)$ is completely continuous on L^p to L.

Likewise we have the following theorem:

THEOREM 72. *If for almost all s in $(0, 1)$ the sequence $a_n(s)$ is in $l^{p'}$ and*

$$U(x) = \sum_{n=1}^{\infty} \xi_n a_n(s) = \psi(s)$$

is in $L(0, 1)$ for every vector $x=\{\xi_n\}$ in l^p, then the operation $U(x)$ on l^p to L is completely continuous.

As instances of Theorem 32 we have, if T is a class with one element and we take first $Y=l^{p'}$, $\Gamma=l^p$ and then $Y=L^{p'}$, $\Gamma=L^p$, the following two theorems:

THEOREM 73. *Every linear operator from l^p, $(p>1)$, to l is completely continuous.*

THEOREM 74. *Let $a_n(s)$ be in $L^{p'}(0, 1)$, $(n=1, 2, \cdots)$, and suppose that*

$$U(\phi) = \int_0^1 a_n(s)\phi(s)ds$$

is in l for every ϕ in L^p. Then $U(\phi)$ is completely continuous.

As a matter of fact *any continuous linear operator from $\overline{Y}$ to l is completely continuous provided $Y=\overline{Y}$.* This fact has been proved by Pettis [25] and follows also from Theorem 30 together with the remarks following Theorem 32.

THEOREM 75. *If $\{y_n\}$ is a sequence of points in a Banach space Y such that $\sum_n \xi_n y_n$ converges for every $x=\{\xi_n\}$ in m, then the operation*

$$U(x) = \sum_{n=1}^{\infty} \xi_n y_n$$

*on m to Y is completely continuous.**

By hypothesis all partial sums of $\sum_n y_n$ are convergent; hence by a result of Orlicz [23] the series $\sum_n y_n$ is unconditionally convergent. The operation $V(\gamma)=\{\gamma y_n\}$ is therefore a linear operation on Y to l. By Theorem 32, V is completely continuous, and since U is the adjoint of V, it follows from a result of Schauder [33] that U is completely continuous also.

If we had been working in the complex domain instead of the real domain, the following theorem would be an immediate corollary of the definition of the integral. As it is, a few words of explanation are necessary. In this final theorem Y is a complex Banach space, that is, a complete normed vector space satisfying all the postulates for a Banach space with the real number system replaced by the complex number system (see Wiener [38]). Let $\overline{Y}$ be the space of all complex valued continuous linear functionals on Y. Then upon placing

$$\|\gamma\| = \sup_{\|y\|=1} |\gamma y|$$

for γ in $\overline{Y}$, the space $\overline{Y}$ is also a complex Banach space.

The space Y may be regarded as a Banach space; thus ([1], p. 55) for each y_0 there is a real function μ such that

$$\mu(y_0) = \|y_0\|, \qquad \sup_{\|y\|=1} |\mu y| = 1,$$

$$\mu(c_1y_1 + c_2y_2) = c_1\mu(y_1) + c_2\mu(y_2), \qquad c_1, c_2 \text{ real}.$$

The function μ is not in $\overline{Y}$ however because $\mu cy \neq c\mu y$ for c complex. However by defining $\mu'(y)=-\mu(iy)$, the function

$$\gamma y = [1/(2)^{1/2}][\mu(y) + i\mu'(y)]$$

is continuous and linear, and $\gamma cy = c\gamma y$ for complex values of c. Thus γ is in $\overline{Y}$. It is readily shown that

$$\|\gamma\| \leqq 1, \qquad |\gamma y_0| \geqq \|y_0\|/(2)^{1/2},$$

* This theorem was proved for the case $Y=l$ by Littlewood [19].

so that

$$\sup_{\|\gamma\|=1} |\gamma y| \geqq \|y\|/(2)^{1/2}, \qquad y \in Y.$$

A *determining manifold* $\Gamma \subset \overline{Y}$ is defined to be a closed linear manifold in $\overline{Y}$ such that

$$\sup_{\|\gamma\|=1} |\gamma y| \geqq M\|y\|, \qquad y \in Y,$$

where γ is restricted to be in Γ and M is some positive constant independent of y.

It is readily shown that a function $y(t)$ on an arbitrary range T is bounded if $\gamma y(t)$ is bounded for every γ in a determining manifold $\Gamma \subset \overline{Y}$. For then Γ is the sum of the closed sets

$$\Gamma_n = \Gamma[|\gamma y(t)| \leqq m, t \in T],$$

and by the Baire category theorem there is a sphere $S(\gamma_0, r)$ of Γ contained in Γ_{n_0} for some integer n_0. Thus if $\|\gamma\| < r$,

$$|\gamma y(t)| \leqq |(\gamma - \gamma_0)y(t)| + |\gamma_0 y(t)| \leqq 2n_0, \qquad \text{on } T,$$

so that, for any γ in Γ with $\gamma \leqq 1$, $|\gamma y(t)| \leqq 2n_0 r$ on T and thus

$$\|y(t)\| \leqq 2n_0 r/M, \qquad \text{on } T.$$

In terms of this notation we have the following theorem:

THEOREM 76. *Suppose D is a simply connected open set in the complex plane, Y is an arbitrary complex Banach space, and $f(z)$ on D to Y is such that $\gamma f(z)$ is analytic for every γ in Γ. Then $\gamma f(z)$ is analytic for every γ in $\overline{Y}$ and**

(i) $\int_C f(z)dz = 0$ *for any rectifiable curve C in D;*

(ii) $f(z) = (1/2\pi i)\int_C [f(\zeta)/(\zeta - z)]d\zeta$ *if C contains z;*

(iii) $f(z)$ *has strong derivatives of all orders and*

(iv) $f^{(n)}(z) = (n!/2\pi i)\int_C [f(\zeta)/(\zeta - z)^{n+1}]d\zeta$ *if C contains z;*

(v) *the Taylor expansion*

$$\sum_{n=0}^{\infty} \frac{(z-\zeta)^n}{n!} f^{(n)}(z)$$

converges uniformly to $f(z)$ for z in any circle $|z - \zeta| \leqq r$ inside D.

All of the above integrals can be taken in the Riemann sense.

To prove (ii) note first that the boundedness of $\gamma f(z)$ on C for every γ in Γ implies the boundedness of $\|f(z)\|$ on C. Now from the formula

* If the boundary of D is a rectifiable Jordan curve, and $\gamma f(z)$ is continuous in the closed domain $\overline{D}$ for each γ in Γ, then the curve C, appearing in the various integrals, may be taken as the bounding curve.

$$\gamma[f(z+h)-f(z)]=\frac{1}{2\pi i}\int_C \gamma f(\zeta)\left[\frac{1}{\zeta-(z+h)}-\frac{1}{\zeta-z}\right]d\zeta$$

and the boundedness of $\|f(z)\|$ on C it follows that

$$\lim_{h=0}\gamma[f(z+h)-f(z)]=0$$

uniformly for $\|\gamma\|\leqq 1$, that is, $\lim_{h=0} f(z+h)=f(z)$. Thus the integral

$$g(z)=\frac{1}{2\pi i}\int_C \frac{f(\zeta)}{\zeta-z}\,d\zeta$$

exists in the Riemann sense. Since $\gamma f(z)=\gamma g(z)$ for every γ in a determining manifold, we have

$$f(z)=\frac{1}{2\pi i}\int_C \frac{f(\zeta)}{\zeta-z}\,d\zeta.$$

The remaining conclusions of the theorem follow in the usual manner from the above formula.

Bibliography

1. S. Banach, *Théorie des Opérations Linéaires*, Warsaw, 1932.

2. S. Banach, *Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales*, Fundamenta Mathematicae, vol. 3 (1922), pp. 133–181; in particular pp. 157–160.

3. G. Birkhoff, *Integration of functions with values in a Banach space*, these Transactions, vol. 38 (1935), pp. 357–378.

4. S. Bochner, *Integration von Funktionen deren Werte die Elemente eines Vektorräumes sind*, Fundamenta Mathematicae, vol. 20 (1933), pp. 262–276.

5. L. W. Cohen and N. Dunford, *Linear operations on sequence spaces*, Duke Mathematical Journal, vol. 3 (1937), pp. 689–701.

6. N. Dunford, *Integration in general analysis*, these Transactions, vol. 37 (1935), pp. 441–453; corrections, ibid. vol. 38 (1936), pp. 600–601.

7. G. Fichtenholz and L. Kantorovitch, *Sur les opérations dans l'espace des fonctions bornées*, Studia Mathematicae, vol. 5 (1934), pp. 69–99.

8. M. Fréchet, *Sur l'intégrale d'une fonctionelle étendue à un ensemble abstrait*. Bulletin de la Société Mathématique de France, vol. 43 (1915), pp. 249–267.

9. I. Gelfand, *Sur un lemme de la théorie des espaces linéaires*, Communications de l'Institut des Sciences Mathématique et Mécaniques de l'Université de Kharkoff et la Société Mathématique de Kharkoff, (4), vol. 13, pp. 35–40.

10. L. M. Graves, *Riemann integration and Taylor's theorem in general analysis*, these Transactions, vol. 29 (1927), pp. 163–177.

11. H. Hahn, *Über Folgen linearer Operationen*, Monatshefte für Mathematik und Physik, vol. 32 (1922), pp. 1–88.

12. H. Hahn, *Über die Darstellung gegebener Funktionen durch singuläre Integrale*, II. Denkschriften der Kaiserlichen Akademie in Wien, Mathematisch-Naturwissenschaftliche Klasse, vol. 93 (1916), pp. 657–692; in particular pp. 678–679.

13. T. H. Hildebrandt, *On uniform limitedness of sets of functional operations*, Bulletin of the American Mathematical Society, vol. 29 (1923), pp. 309–315.

14. T. H. Hildebrandt, *Lebesgue integration in general analysis*, Bulletin of the American Mathematical Society, vol. 33 (1927), p. 646.

15. T. H. Hildebrandt, *On bounded linear functional operations*, these Transactions, vol. 36 (1934), pp. 868–875.

16. T. H. Hildebrandt, *Linear operations on functions of bounded variation*, Bulletin of the American Mathematical Society, vol. 44 (1938), pp. 75–76.

17. M. Krein, *Sur quelque questions de la géométrie des ensembles convexes situés dans un espace linéaire normé et complet*, Comptes Rendus de l'Académie des Sciences de l'URSS, vol. 14 (1937), pp. 5–7.

18. H. Lebesgue, *Sur les intégrales singulières*, Annales de Toulouse, (3), vol. 1 (1909), pp. 25–117; in particular pp. 60–62.

19. J. E. Littlewood, *On bounded bilinear forms in an infinite number of variables*, Quarterly Journal of Mathematics, Oxford, vol. 1 (1930), pp. 164–174.

20. Z. Łomnicki and S. Ulam, *Sur la théorie de la mesure dans les espaces combinatoires et son application au calcul des probabilités*, Fundamenta Mathematicae, vol. 22 (1934), pp. 237–278.

21. J. von Neumann, *Über adjungierte Funktionaloperatoren*, Annals of Mathematics, (2), vol. 33 (1932), pp. 294–310.

22. O. Nikodym, *Sur une généralisation des intégrales de M. J. Radon*, Fundamenta Mathematicae, vol. 15 (1930), pp. 131–179.

23. W. Orlicz, *Beiträge zur Theorie der Orthogonalentwicklungen*, II, Studia Mathematica, vol. 1 (1929), pp. 241–255; in particular pp. 244–247.

24. B. J. Pettis, *On integration in vector spaces*, these Transactions vol. 44 (1938), pp. 277–304.

25. B. J. Pettis, *A note on regular Banach spaces*, Bulletin of the American Mathematical Society, vol. 44 (1938), pp. 420–428.

26. H. R. Pitt, *A note on bilinear forms*, Journal of the London Mathematical Society, vol. 11 (1936), pp. 174–180.

27. S. Pollard, *The Stieltjes integral and its generalizations*, Quarterly Journal of Mathematics, vol. 49 (1920), pp. 73–138.

28. J. Radon, *Theorie und Andwendungen der absolut additiven Mengenfunktionen*, Sitzungsberichte der Akademie der Wissenschaften, Vienna, Class IIa, vol. 122 (1913), pp. 1295–1438.

29. J. Ridder, *Integration in abstrakten Räumen*, Fundamenta Mathematicae, vol. 24 (1935), pp. 72–117.

30. S. Saks, *Théorie de l'Intégrale*, Warsaw, 1933.

31. S. Saks, *Addition to the note on some functionals*, these Transactions, vol. 35 (1933), p. 967.

32. S. Saks and J. D. Tamarkin, *On a theorem of Hahn-Steinhaus*, Annals of Mathematics, (2), vol. 34 (1933), pp. 595–601.

33. J. Schauder, *Über lineare, vollstetige Funktionaloperationen*, Studia Mathematica, vol. 2 (1930), pp. 183–196; in particular Theorem 1.

34. M. H. Stone, *Linear Transformations in Hilbert Space and their Applications to Analysis*, American Mathematical Society Colloquium Publications, vol. 15, New York, 1932.

35. M. H. Stone and J. D. Tamarkin, *Elementary proofs of some known theorems of the theory of complex euclidean spaces*, Duke Mathematical Journal, vol. 3 (1937), pp. 294–302.

36. O. Toeplitz, *Über allgemeine lineare Mittelbildungen*, Prace Mathematyczno-Fizycne, vol. 22 (1911), pp. 113–119.

37. A. Tulajkov, *Zur Kompaktheit im Raum L_p für $p=1$*, Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, 1933, pp. 169–170.

38. N. Wiener, *Note on a paper of M. Banach*, Fundamenta Mathematicae, vol. 4 (1923), pp. 136–143.

YALE UNIVERSITY,
NEW HAVEN, CONN.

GROUPS WITH ABELIAN CENTRAL QUOTIENT GROUP†

BY

REINHOLD BAER

In recent years much progress has been made in the study of groups whose central quotient groups are abelian.‡ Such a group is an extension of an abelian group by an abelian group, and the usual method of solving the implied extension problem may be described as follows: If G is such a group, then a maximum abelian subgroup V of G is chosen. V contains the central of G, and G/V represents therefore exactly an abelian group of automorphisms of the abelian group V. Thus it is possible to apply the results of the theory of automorphisms of abelian groups. This method is rather powerful and yields very interesting results. On the other hand it is not restricted to this class of groups and may in fact be applied to all groups with abelian commutator groups.§ Finally it has to be mentioned that this method is not an invariant one, since the maximum abelian subgroups are in no sense uniquely determined.

In this paper another method will be indicated. If G is a group whose central quotient group is abelian, then preference is given to a subgroup S which is situated between the central and the commutator group of G. This subgroup S will be left indeterminate as long as possible. But as soon as the final results are reached, either the central or the commutator group of G will take the place of S according to which of these choices will give better results.

The extension problem presents itself now in the following form: To characterize those groups whose central contains a given abelian group S and whose quotient group (mod S) is isomorphic to a preassigned abelian group G^*. Each group with these properties induces certain invariant relations between the given abelian groups G^* and S, and these invariants turn out to be characteristic invariants, provided G^* is a direct product of (a finite or infinite number of finite and infinite) cyclic groups. This last hypothesis will be the only restriction of generality imposed on the investigated groups, so that the finite groups with abelian central quotient groups are included in our treatment.

† Presented to the Society, September 7, 1937; received by the editors September 20, 1937.

‡ Cf., for example, H. R. Brahana, American Journal of Mathematics, vol. 57 (1935), pp. 645–667, and Duke Mathematical Journal, vol. 1 (1935), pp. 185–197; also C. Hopkins, these Transactions, vol. 37 (1935), pp. 161–195; vol. 41 (1937), pp. 287–313.

§ K. Taketa, Japanese Journal of Mathematics, vol. 13 (1937), pp. 129–232.

1. **Basic concepts and formulas.** If G is any group, then $Z(G)$ denotes the central, and $C(G)$ the commutator group of G. If n is any positive integer, then G_n is generated by the elements x which satisfy $x^n=1$, and G^n is generated by the elements of the form x^n for x in G. (If G is an abelian group, then G_n consists of the elements x with $x^n=1$, and G^n consists of the elements x^n for x in G.) Finally let $(x, y)=xyx^{-1}y^{-1}$.

(1.1) *If G and S are groups such that $C(G)\leqq S\leqq Z(G)$, then the operation $(S<G; x^*, y^*)=x^*\circ y^*$, which is defined by $Sx\circ Sy=(x, y)$, obeys the following rules*:

(1) *$x^*\circ y^*$ is, for every x^*, y^* in $G^*=G/S$, a uniquely determined element in S.*

(2) $1=x^*\circ x^*=(x^*\circ y^*)(y^*\circ x^*)$.

(3) $x^*\circ y^*z^*=(x^*\circ y^*)(x^*\circ z^*)$.

(4) *The element z in G belongs to $Z(G)$ if, and only if, $Sz\circ x^*=1$ for every x^* in G^*.*

Proof. The statements (1), (2), and (4) are obvious. (3) may be proved as follows:

$$\begin{aligned} Sx\circ SySz &= xyzx^{-1}z^{-1}y^{-1} = xyx^{-1}xzx^{-1}z^{-1}y^{-1} \\ &= (x, y)(x, z), \end{aligned}$$

since every commutator is an element of the central.

An obvious consequence of assertion (1.1), (4) is the following proposition:

(1.2) *Suppose that $C(G)\leqq S\leqq Z(G)$. Then $S=Z(G)$ if, and only if, 1 is the only element w^* in G/S such that $w^*\circ x^*=1$ for every element x^* in G/S.*

(1.3) *If x and y are elements of the group G with abelian central quotient group, then*

$$(xy)^i = (y, x)^{i(i-1)2^{-1}}x^iy^i$$

for every positive integer i.

Proof. This statement is certainly true for $i=1$. If $1<i$, then

$$\begin{aligned} x^{i-1}y^{i-1}xy &= x^ix^{-1}y^{i-1}xy^{1-i}y^i = x^i(x^{-1}, y^{i-1})y^i \\ &= (y, x)^{i-1}x^iy^i, \end{aligned}$$

since the commutators are elements of the central, and since (2) and (3) of (1.1) may be applied. If (1.3) holds true for $i-1$, then

$$(xy)^i = (y, x)^{(i-1)(i-2)2^{-1}}x^{i-1}y^{i-1}xy = (y, x)^{i(i-1)2^{-1}}x^iy^i,$$

and thus (1.3) holds true for every i.

(1.4) *Suppose that* $C(G) \leqq S \leqq Z(G)$ *and that* x *and* y *are elements of* G *such that* Sx *and* Sy *are both contained in* $(G/S)_n$.

(a) *If at least one of the elements* Sx *and* Sy *is contained in* $(G/S)^2$ *(this would certainly be the case, if* n *is odd), then* $(xy)^n = x^n y^n$.

(b) *If* $n = 2m$ *is an even integer, then*

$$(xy)^{2m} = (Sx \circ Sy)^m x^{2m} y^{2m} \quad \text{and} \quad ((Sx \circ Sy)^m)^2 = 1.$$

Proof. It follows from (1.3) that

$$(xy)^n = (Sy \circ Sx)^{n(n-1)2^{-1}} x^n y^n;$$

and the statement (b) is now a consequence of (1.1). In order to prove (a), assume that Sy belongs to $(G.S)^2$. Then $Sy = Sz^2$ for some z in G and

$$(Sy \circ Sx)^{n(n-1)2^{-1}} = (Sz \circ Sx^n)^{n-1} = 1$$

by (1.1). This completes the proof.

(1.5) *If* $C(G) \leqq S \leqq Z(G)$, *and if the function* $P(n, x^*) = P(S < G; n, x^*)$ *is defined by* $P(n, Sx) = S^n x^n$, *then the following statements are true*:

(1) $P(n, x^*)$ *is, for every* x^* *in* $(G/S)_n$, *a uniquely determined element in* S/S^n.

(2′) *If* x^* *and* y^* *are both elements of* $(G/S)_n$ *and at least one of them is an element of* $(G/S)^2$ *(for example, if* n *is odd), then* $P(n, x^*y^*) = P(n, x^*)P(n, y^*)$.

(2″) *If* x^* *and* y^* *are both elements of* $(G/S)_{2m}$, *then*

$$P(2m, x^*y^*) = (x^* \circ y^*)^m P(2m, x^*) P(2m, y^*),$$

and $(x^* \circ y^*)^m$ *is either the unit or an element of order* 2.

(3) $P(nm, x^*) = P(n, x^*)^m$ *for positive* m *and* x^* *in* $(G/S)_n$; *and* $P(n, x^{*m}) = S^n P(nm, x^*)$ *for* x^* *in* $(G/S)_{nm}$.

Proof. (1), (2′), and (2″) are consequences of (1.4). If Sx belongs to $(G/S)_n$ and m is a positive integer, then Sx belongs to $(G/S)_{nm}$ and $P(n, Sx)^m = (S^n x^n)^m = S^{nm} x^{nm} = P(nm, Sx)$. If Sx belongs to $(G/S)_{nm}$, then Sx^m belongs to $(G/S)_n$, and $P(n, Sx^m) = S^n x^{nm} = S^n S^{nm} x^{nm} = S^n P(nm, Sx)$.

2. **Existence theorems.** We prove the following theorem:

THEOREM 2.1. *Suppose that* S *and* G^* *are abelian groups, that* $x^* \circ y^*$ *is, for every* x^* *and* y^* *in* G^*, *an element in* S, *and that* $P(n, x^*)$ *is an element in* S/S^n *for every positive integer* n *(if* n *is such that elements of order* n *exist in* G^**) and for every* x^* *in* G_n^*.

There exists a group G *such that* $C(G) \leqq S \leqq Z(G)$ *and an isomorphism* γ *of* G/S *upon the whole group* G^* *such that* $(Sx)^\gamma \circ (Sy)^\gamma = (x, y)$ *for* x *and* y *in* G *and* $P(n, (Sx)^\gamma) = S^n x^n$ *for* Sx *in* $(G/S)_n$ *(provided there exist elements of order* n *in* G^**) if, and only if, the following conditions are satisfied*:

(1) $1=x^*\circ x^*=(x^*\circ y^*)(y^*\circ x^*)$, $x^*\circ y^*z^*=(x^*\circ y^*)(x^*\circ z^*)$.

(2) $P(n,\ x^*y^*)=P(n,\ x^*)P(n,\ y^*)$, *if the* n *is odd; and* $P(2m,\ x^*y^*)=(x^*\circ y^*)^m P(2m, x^*)P(2m, y^*)$.

(3) $P(nm, x_m^*)=P(n, x^*)^m$, *if* x^* *is an element of* G_n^*. $P(n, x^*)=S^n P(nm, x^*)$, *if* x^* *is an element of* G_{mn}^*.

Proof. The necessity of the conditions (1) to (3) is a consequence of (1.1) and (1.5). Suppose now that conditions (1) to (3) are satisfied. There exist an ordinal number t and, for every ordinal number v with $0\leqq v\leqq t$, a group $G^*(v)$ with the following properties: $G^*(0)=1$, $G^*(v)<G^*(v')$ for $v<v'$, $G^*(t)=G^*$; $G^*(v+1)/G^*(v)$ is a cyclic group whose order is either infinite or a prime number; and if v is a limit ordinal, then $G^*(v)$ is the join of the groups $G^*(u)$ for $u<v$.

By complete (transfinite) induction with regard to v, groups $G(v)$ and isomorphisms $\gamma(v)$ will be defined which obey the following rules:

(i) $C(G(v))\leqq S\leqq Z(G(v))$.

(ii) $G(v)<G(v')$ for $0\leqq v<v'\leqq t$.

(iii) $\gamma(v)$ is an isomorphism of $G(v)/S$ upon the whole group $G^*(v)$.

(iv) If x is an element in $G(v)$ and $v<v'$, then $(Sx)^{\gamma(v)}=(Sx)^{\gamma(v')}$.

(v) If x and y are elements in $G(v)$, then $(x, y)=(Sx)^{\gamma(v)}\circ(Sy)^{\gamma(v)}$.

(vi) If G^* contains elements of the finite order n, and if Sx belongs to $(G(v)/S)_n$, then $S^n x^n=P(n, (Sx)^{\gamma(v)})$.

The choices $G(0)=S$ and $\gamma(0)=1$ are in accordance with these rules, since $G^*(0)=1$. It may therefore be assumed that for every $u<v$ a group $G(u)$ and an isomorphism $\gamma(u)$ have been defined which satisfy (i) to (vi).

Case 1. $v=w+1$ is not a limit ordinal. $G^*(w+1)/G^*(w)$ is a cyclic group. It is either infinite or its order is a prime number p. Denote by b^* an element of $G^*(v)$ generating $G^*(v)$ (mod $G^*(w)$). If $G^*(w)b^*$ contains elements of finite order, then b^* may be chosen as an element of minimum order in its class $G^*(w)b^*$. If $G^*(w+1)/G^*(w)$ is of finite order p, then let c be an element in $G(w)$ such that $(Sc)^{\gamma(w)}=b^{*p}$.

In order to define $G(v)$ and $\gamma(v)$ we require first the following results:

(2.11) *An automorphism* β *of* $G(w)$ *is defined by* $x^\beta=(b^*\circ(Sx)^{\gamma(w)})x$. *All the elements of* S *are fixed elements of* β. *If* $G^*(v)/G^*(w)$ *is of finite order* p, *then* c *is a fixed element of* β *and the automorphism* $\psi=\beta^p$ *is the inner automorphism of* $G(w)$ *induced by* c.

By its definition x^β is, for every x in $G(w)$, a uniquely determined element of $G(w)$. If x is an element in S, then $Sx=S$ and $x^\beta=x$ by (1). If x and y are elements of $G(w)$, then $(xy)^\beta=(b^*\circ(Sxy)^{\gamma(w)})xy=(b^*\circ(Sx)^{\gamma(w)})(b^*\circ(Sy)^{\gamma(w)})xy=x^\beta y^\beta$ by (1) and (i). If $x^\beta=1$, then x is an element of S, consequently $x=1$.

From these facts it follows that β is a (proper) automorphism of $G(w)$.

Suppose now that $G^*(v)/G^*(w)$ is of finite order p. Then $c^\beta = (b^* \circ (Sc)^{\gamma(w)})c = (b^* \circ b^{*p})c = c$ by (1). Since the elements of S are fixed elements of the automorphism β,

$$x^\psi = (b^* \circ (Sx)^{\gamma(w)})^p x = (b^{*p} \circ (Sx)^{\gamma(w)})x$$

by (1), and

$$x^\psi = ((Sc)^{\gamma(w)} \circ (Sx)^{\gamma(w)})x = cxc^{-1}$$

by (v). This completes the proof of (2.11).

(2.12) *There exists a group $G(v)$ with the following properties*:

(a) *$G(w)$ is a normal subgroup of $G(v)$, and $G(v)/G(w)$ is a cyclic group of the same order as $G^*(v)/G^*(w)$.*

(b) *$G(v)$ contains an element b which generates $G(v)$* (mod $G(w)$) *and satisfies the relation $bxb^{-1} = x^\beta$ for every x in $G(w)$.*

(c) *If $G^*(v)/G^*(w)$ is of finite order p, then b^p is an element of $G(w)$, and $Sc = Sb^p$.*

(d) *If b^* is an element of finite order pn, then $S^{pn}b^{pn} = P(pn, b^*)$.*

To prove this note first that $G^*(v)/G^*(w)$ is of finite order p if b^* is of finite order. The order of b^*, if finite, is therefore a multiple pn of p. It follows from (3) and (vi) that there exists an element c', satisfying $Sc = Sc'$ and $P(pn, b^*) = S^{pn}c'^n$. Now it follows from (2.11) and general theorems in the theory of extensions of groups† that the group $G(v)$ which is generated by $G(w)$ and an element b subject to the relations $bxb^{-1} = x^\beta$ for x in $G(w)$ and $b^p = c$, if b^* is of infinite order and $G^*(v)/G^*(w)$ of finite order p, or $b^p = c'$, if b^* is of finite order pn and $G^*(v)/G^*(w)$ of finite order p, satisfies the conditions (a) to (d).

(2.13) *There exists one and only one isomorphism $\gamma(v)$ of $G(v)/S$ upon $G^*(v)$ which satisfies*

$$(Sb)^{\gamma(v)} = b^*, \qquad (Sx)^{\gamma(v)} = (Sx)^{\gamma(w)}$$

for x in $G(w)$.

This is an obvious consequence of the following facts: β induces in S and in $G(v)/S$ the identical automorphism; if $G^*(v)/G^*(w)$ is of finite order p, then $b^{*p} = (Sc)^{\gamma(w)} = (Sc')^{\gamma(w)}$.

That these definitions of $G(v)$ and $\gamma(v)$ are in accordance with (i) to (iv) is clear. Any pair of elements in $G(v)$ has the form xb^i, yb^j with x, y in $G(w)$. Hence, by (1), (i), (v) is a consequence of

† A. M. Turing, *The extension of groups*, Compositio Mathematica, vol. 5 (1938), pp. 557–567.

$$(xb^i, yb^j) = xb^iyb^jb^{-i}x^{-1}b^{-j}y^{-1} = xy^{\beta^i}(x^{-1})^{\beta^j}y^{-1}$$
$$= (b^{*i} \circ (Sy)^{\gamma(w)})((Sx)^{\gamma(w)} \circ b^{*j})(x, y)$$
$$= (Sxb^i)^{\gamma(v)} \circ (Syb^j)^{\gamma(v)}.$$

Suppose that G^* contains elements of order m and that xb^i is an element in $G(v)$ such that $(xb^i)^m$ is an element of S. Then it follows from (i) and (1.3) that

$$(xb^i)^m = ((Sx)^{\gamma(w)} \circ b^*)^{-im(m-1)2^{-1}}x^mb^{im};$$

consequently x^mb^{im} is an element of S. Therefore $i=0$, if $G^*(v)/G^*(w)$ is infinite, and $S^mx^m = P(m(Sx)^{\gamma(w)})$ by (vi) (which condition is satisfied in $G(w)$). If $G^*(v)/G^*(w)$ is of finite order p, then it may be assumed without loss in generality that $0 \leqq i < p$. If $i=0$, then the above argument may be applied; if $i \neq 0$, then i and p are relatively prime. Since b^{im} is an element of $G(w)$, this implies that m is a multiple of p. Since i is relatively prime to p, and since b^i and xb^i are elements of the same class (mod $G(w)$), it follows that the order of b^{*i} is finite and a divisor of the order of $(Sxb^i)^{\gamma(v)}$. Hence pn is a divisor of m, and b^{im}, as well as x^m, is an element of S. Consequently

$$S^m(xb^i)^m = ((Sx)^{\gamma(w)} \circ b^*)^{-im(m-1)2^{-1}}P(m, (Sx)^{\gamma(w)})P(m, b^{*i})$$

by (vi) applied on $G(w)$, by (2.12), (d) and by condition (3); and

$$S^m(xb^i)^m = P(m, (Sx)^{\gamma(w)}b^{*i}) = P(m, (Sxb^i)^{\gamma(v)})$$

by (2) and (2.13). Thus the condition (vi) is satisfied by $G(v)$ and $\gamma(v)$.

Case 2. v is a limit ordinal. Let $G(v)$ be the uniquely determined group which contains all the groups $G(u)$ for $u < v$ and is just their join, and let $\gamma(v)$ be the uniquely determined isomorphism of $G(v)/S$ upon $G^*(v)$ which coincides on every $G(u)/S$ for $u < v$ with $\gamma(v)$. That the conditions (i) to (vi) are satisfied is clear.

Thus $G(v)$, $\gamma(v)$ are defined for every $0 \leqq v \leqq t$. Since $G^*(t) = G^*$, it follows that the group $G(t) = G$ and the isomorphism $\gamma(t) = \gamma$ meet all the requirements of the theorem. This completes the proof.

COROLLARY 2.2. *Suppose that S and G^* are abelian groups and that $x^* \circ y^*$ is, for every x^* and y^* in G^*, an element in S. Then there exists a group G such that $C(G) \leqq S \leqq Z(G)$ and an isomorphism γ of G/S upon the whole group G^* such that*

$$(Sx)^\gamma \circ (Sy)^\gamma = (x, y)$$

for x and y in G if, and only if,

$$1 = x^* \circ x^* = (x^* \circ y^*)(y^* \circ x^*), \qquad x^* \circ y^*z^* = (x^* \circ y^*)(x^* \circ z^*).$$

Proof. The necessity of the condition is a consequence of (1.1). Suppose now that the condition is satisfied. Denote by $F(G^*)$ the subgroup of all the elements of finite order in G^*. Then there exists a basis B^* of $F(G^*)$ (mod $F(G^*)$)† which contains, for every positive integer i, a basis B_i^* of $G_{2^i}^*$ modulo the cross cut of G^{*2} and $G_{2^i}^*$.‡ (B_i is the cross cut of B^* and $G_{2^i}^*$.) If n is an odd integer and m a non-negative integer, then every element x^* in $G_{n2^m}^*$ may be represented in one and only one way in the form

$$x^* \equiv \prod_{b^* \in B_m^*} b^{*h(x^*, b^*)} \pmod{G^{*2}},$$

$h(x^*, b^*) = 0$ or 1, and $h(x^*, b^*) = 0$ for almost every b^*. If $b_1^*, \cdots, b_k^*$ are exactly those elements b^* in B_m^* such that $h(x^*, b_j^*) = 1$, then put

$$P(n2^m, x^*) = S^{n2^m} \prod_{j<i} (b_j^* \circ b_i^*)^{2^{m-1}}.$$

It is now fairly obvious that the functions $x^* \circ y^*$ and $P(r, x^*)$ satisfy the conditions (1) to (3) of the Theorem 2.1, and the existence of a group G and of an isomorphism γ, meeting the requirements of the Corollary 2.2, is a consequence of Theorem 2.1.

COROLLARY 2.3. *If Z and G are abelian groups, then the necessary and sufficient condition for the existence of a group whose central is Z and whose central quotient group is G is the existence of an operation $x \circ y$ of the elements x and y in G with values in Z, satisfying the conditions:*

(1) *$x \circ y$ is for every x and y in G an element in Z.*
(2) *$1 = x \circ x = (x \circ y)(y \circ x)$.*
(3) *$x \circ yz = (x \circ y)(x \circ z)$.*
(4) *$w \circ x = 1$ for every x in G if, and only if, $w = 1$.*

This corollary is a consequence of Corollary 2.2 and of (1.2).

3. **Factor sets.** Suppose that $C(G) \leqq S \leqq Z(G)$ and that $G^* = G/S$ is a direct product of cyclic groups. Let B^* be a basis of G^*, and let B be some set of representatives in G of the classes in B^*. The elements in B may be ordered in some way. It is important to note that the following formulas depend on the way in which B has been ordered.

If x^* is any element of finite order in G^*, then let $n(x^*)$ be its order. Now every element x in G may be represented in one and only one way as

$$x = s(x) \prod_{b \in B} b^{n(x,b)},$$

† K. Taketa, Japanese Journal of Mathematics, vol. 13 (1937), pp. 129–232.

‡ Cf., for example, R. Baer, American Journal of Mathematics, vol. 59 (1937), pp. 99–117, particularly §1.

where the factors of the product are ordered in the same way as B, almost every $n(x, b)=0$, where $0 \leqq n(x, b) < n(Sb)$ (if Sb is an element of finite order), and $s(x)$ is an element in S.

If x and y are two elements in G, then integers $h(x, y; b)$, which may only be 0 or 1, are uniquely determined by the equation

$$n(xy, b) + h(x, y; b)n(Sb) = n(x, b) + n(y, b)$$

if Sb is of finite order, and

$$n(xy, b) = n(x, b) + n(y, b)$$

if Sb is of infinite order. Put furthermore

$$a(x, y) = \prod' b^{n(Sb)},$$

where $\prod'$ indicates a product taken over all those b in B such that Sb is of finite order and $h(x, y; b)=1$; and put

$$c(x, y) = \prod_{b<b'} (Sb' \circ Sb)^{n(x,b')n(y,b)},$$

where $Sb \circ Sb' = (b, b')$. Then it follows from (1.1) that

$$s(xy) = s(x)s(y)a(x, y)c(x, y).$$

Note that $a(x, y)=a(Sx, Sy)=a(Sy, Sx)$ and $c(x, y)=c(Sx, Sy)$. The product $a(x^*, y^*)c(x^*, y^*)$ is, in the usual terminology,† a factor set of the extension G of S by G^* satisfying $S \leqq Z(G)$.

4. **Existence of transformations.** Suppose that S, G and T, H are groups with the property that $C(G) \leqq S \leqq Z(G)$ and $C(H) \leqq T \leqq Z(H)$. Then ϕ is called an *S-T-transformation of G into H*, if ϕ obeys the following postulates:

(1) *x^ϕ is, for every x in G, a uniquely determined element in H.*

(2) *$(sx)^\phi = s^\phi x^\phi$ for s in S and x in G.*

(3) *ϕ maps S exactly upon the whole group T.*

(4) *ϕ induces in $G^*=G/S$ a homomorphism λ upon the whole group $H^*=H/T$.*

(5) *If n is a positive integer, and if x and y are elements in G and x^n, y^n elements in S, then $((xy)^n)^\phi = ((xy)^\phi)^n = (x^\phi y^\phi)^n$.*

It follows from (2) and (3) that the S-T-transformation ϕ induces in S a homomorphism σ upon T, and it follows from (1) to (4) that every element in H is, under ϕ, the picture of some element in G. Using the conditions (1) to (4) we may analyze the condition (5) in the following way. By choosing $y=1$, (5) specializes to the statement:

† Cf., for example, R. Baer, Mathematische Zeitschrift, vol. 38 (1934), pp. 375–416.

(5′) If n is a positive integer, x an element in G, and x^n an element in S, then $(x^n)^\phi = (x^\phi)^n$.

Since ϕ induces a homomorphism of S upon the whole group T, it follows that $(S^n)^\phi = T^n$, consequently (5′) implies the condition:

CONDITION 4.a. $P(S<G; n, x^*)^\sigma = P(T<H; n, x^{*\lambda})$ *for* x^* *in* G_n^*.

If (1) to (4) and (5′) hold, and if Sx and Sy are elements in G_n^*, then

$$\begin{aligned}((xy)^\phi)^n &= ((xy)^n)^\phi = ((y, x)^{n(n-1)2^{-1}}x^n y^n)^\phi \\ &= ((y, x)^{n(n-1)2^{-1}})^\phi (x^n)^\phi (y^n)^\phi, \\ (x^\phi y^\phi)^n &= (y^\phi, x^\phi)^{n(n-1)2^{-1}}(x^\phi)^n (y^\phi)^n \\ &= (y^\phi, x^\phi)^{n(n-1)2^{-1}}(x^n)^\phi (y^n)^\phi.\end{aligned}$$

This follows from (1.3) and the facts that ϕ is a homomorphism of S upon T and of G^* upon H^*, and that x^n, y^n are in S. In view of (1.1) it is fairly obvious that under the assumption of (1) to (4), condition (5) holds if, and only if, condition (5′) and the following condition are satisfied:

CONDITION 4.b. $((S<G; x^*, y^*)^{2^{m-1}})^\sigma = (T<H; x^{*\lambda}, y^{*\lambda})^{2^{m-1}}$ *for* x^* *and* y^* *in* $G_{2^m}^*$.

THEOREM 4.1. *Suppose that* S, G *and* T, H *are groups with the property* $C(G) \leqq S \leqq Z(G)$ *and* $C(H) \leqq T \leqq Z(H)$, *that* $G^* = G/S$ *is a direct product of cyclic groups, and that conditions* 4.a *and* 4.b *are satisfied by the homomorphism* σ *of* S *upon* T *and the homomorphism* λ *of* G^* *upon* $H^* = H/T$. *Then there exists an S-T-transformation* γ *of* G *into* H *which induces* σ *in* S, *and* λ *in* G^* *and satisfies the condition*:

(6) *If* T' *is the subgroup of* T, *generated by the elements*

$$(T<H; y^{*\lambda}, x^{*\lambda})(S<G, x^*, y^*)^\sigma$$

for x^* *and* y^* *in* G^*, *then* γ *induces a homomorphism of* G *upon* H/T'.

Remark. An obvious consequence of the condition (6) is the following condition:

(6′) *If* $C(G) \leqq U \leqq S$ *and* $C(H) = U^\sigma$, *then* γ *induces a homomorphism of* G/U *upon* H/U^σ.

Proof. Let B be an ordered set in G representing a basis of G/S. If b is any element in B, and if the order of Sb is finite, then it follows from Condition 4.a that there exists an element $f(b)$ in $(Sb)^\lambda$ such that $(b^{n(Sb)})^\sigma = f(b)^{n(Sb)}$, where $n(Sb)$ is the order of Sb.

It has been mentioned in §3 that every element x in G may be represented in one and only one way in the form

$$x = s(x)\prod_b b^{n(x,b)},$$

where $s(x)$ is an element in S, almost every $n(x, b) = 0$, and $0 \leqq n(x, b) < n(Sb)$, if Sb is an element of finite order $n(Sb)$. A function γ may be defined for every x in G by the equation

$$x^\gamma = s(x)^\sigma \prod_{b \in B} f(b)^{n(x,b)},$$

where the factors in the product are ordered in the same way as the elements b in B.

The element x^γ is, for every x in G, uniquely determined in H. If s is an element in S, and x in G, then $s(sx) = ss(x)$ and $n(sx, b) = n(x, b)$; consequently $(sx)^\gamma = s^\sigma x^\gamma = s^\gamma x^\gamma$. This implies also that γ induces σ in S. Since B is a basis of G (mod S), and since $(Sb)^\gamma = Tf(b) = (Sb)^\lambda$, it follows that γ induces λ in G^*.

Suppose now that x is an element in G and n a positive integer such that x^n is an element in S. Then $n(x, b) = 0$ if Sb is of infinite order, and $n(x, b)n = n'(x, b)n(Sb)$ if Sb is of finite order. It follows from (1.3) that

$$x^n = s(x)^n a(x^n) c(x^n),$$

where

$$a(x^n) = \prod_{b \in B} b^{n(x,b)n} = \prod{}'(b^{n(Sb)})^{n'(x,b)}$$

($\prod'$ indicates a product over those b such that Sb is of finite order), and

$$c(x^n) = \prod_{b < b'} (S < G; Sb', Sb)^{n(x,b')n(x,b)n(n-1)2^{-1}}.$$

Computing $a((x^\gamma)^n)$ and $c((x^\gamma)^n)$, accordingly, one finds that

$$a(x^n)^\gamma = a(x^n)^\sigma = \prod{}' f(b)^{n(Sb)n'(x,b)} = a((x^\gamma)^n)$$

and it follows from Condition 4.b that

$$c(x^n)^\gamma = c(x^n)^\sigma = c((x^\gamma)^n).$$

Thus finally

$$(x^n)^\gamma = (x^n)^\sigma = (s(x)^\sigma)^n a((x^\gamma)^n) c((x^\gamma)^n) = (x^\gamma)^n.$$

Hence γ satisfies the conditions (1) to (4) and (5′), and, since γ satisfies Condition 4.b, this implies that γ is an S-T-transformation of G into H.

If x and y are two elements in G, if $a(x, y)$ and $c(x, y)$ are defined as in §3, and if $a(x^\gamma, y^\gamma)$ and $c(x^\gamma, y^\gamma)$ are computed accordingly, then

$$a(x, y)^\gamma = a(x, y)^\sigma = a(x^\gamma, y^\gamma),$$

and

$$(xy)^\gamma = \left(s(x)s(y)a(x, y)c(x, y)\prod_{b\in B} b^{n(xy,b)}\right)^\gamma$$
$$= s(x)^\sigma s(y)^\sigma a(x^\gamma, y^\gamma)c(x, y)^\sigma \prod_{b\in B} f(b)^{n(xy,b)}$$
$$= c(x, y)^\sigma c(x^\gamma, y^\gamma)^{-1}x^\gamma y^\gamma.$$

This shows that γ satisfies (6) also.

LEMMA 4.2. *If $C(G)\leqq S\leqq Z(G)$, if $C(H)\leqq T\leqq Z(H)$, and if ϕ is an S-T-transformation of G into H which induces in S the homomorphism σ and in G/S the homomorphism λ, then a necessary and sufficient condition for ϕ to be a one-one correspondence is that σ and λ are isomorphisms.*

Proof. The necessity of the conditions is clear. If the conditions are satisfied, and if x and y are two elements in G such that $x^\phi = y^\phi$, then $(Sx)^\lambda = (Sy)^\lambda$, and consequently $Sx = Sy$; that is, $x = sy$ for some s in S. Hence $y^\phi = x^\phi = (sy)^\phi = s^\sigma y^\phi$ and $x = y$, since $s^\sigma = 1$ implies $s = 1$.

5. **Homomorphisms and isomorphisms.** We prove the following theorem:

THEOREM 5.1. *Suppose that S, G and T, H are groups with the property $C(G)\leqq S\leqq Z(G)$ and $C(H)\leqq T\leqq Z(H)$, that G/S is a direct product of cyclic groups, that σ is a homomorphism of S upon T and λ a homomorphism of $G^* = G/S$ upon $H^* = H/T$. Then there exists a homomorphism of G into H which induces σ in S and λ in G^* if, and only if,*

(a) $(S<G; x^*, y^*)^\sigma = (T<H; x^{*\lambda}, y^{*\lambda})$ *for x^*, y^* in G^*,*

(b) $P(S<G; n, x^*)^\sigma = P(T<H; n, x^{*\lambda})$ *for x^* in G_n^*.*

A homomorphism of G into H, inducing σ in S and λ in G^*, is an isomorphism if, and only if, σ and λ are isomorphisms.

Proof. The necessity of these conditions is obvious. If, on the other hand, σ and λ satisfy the conditions (a) and (b), then it follows from Theorem 4.1 that there exists an S-T-transformation γ of G into H which induces σ in S and λ in G^* and satisfies the condition (6) of Theorem 4.1. But by condition (a) it follows that the subgroup T' of T, mentioned in (6), consists of the identity only, and γ is consequently a homomorphism. The last statement of Theorem 5.1 is an obvious consequence of Lemma 4.2.

6. **Uniqueness theorems.** We prove the following theorems:

THEOREM 6.1. *Suppose that S, G and T, H are groups with the property $C(G)\leqq S\leqq Z(G)$, $C(H)\leqq T\leqq Z(H)$, and that G/S is a direct product of cyclic groups. Then there exists an isomorphism of G upon H which maps S upon T if, and only if, there exists an isomorphism σ of S upon T and an isomorphism λ of $G^* = G/S$ upon $H^* = H/T$ satisfying conditions* (a) *and* (b) *of Theorem* 5.1.

This is an obvious consequence† of Theorem 5.1.

THEOREM 6.2. *Suppose that $G/Z(G)$ is a direct product of cyclic groups. Then G and the group H are isomorphic if, and only if, there exists an isomorphism σ of $Z(G)$ upon $Z(H)$ and an isomorphism λ of $G/Z(G)$ upon $H/Z(H)$ which satisfy the conditions* (a) *and* (b) *of Theorem* 5.1.

This is a consequence of Theorem 6.1 and the fact that an isomorphism of G upon H maps $Z(G)$ upon $Z(H)$.

Remark 6.3. *If one assumes that G and H are groups with abelian central quotient groups, then Theorem* 6.2 *is transformed into another true theorem, if everywhere the commutator group is substituted for the central.*

Remark 6.4. *If $G/Z(G)$ is abelian and countable and does not contain elements of infinite order, then $G/Z(G)$ is a direct product of cyclic groups.*

Since $G^*=G/Z(G)$ is abelian and does not contain elements of infinite order, it is the direct product of its primary components. If w^* is an element in G^* whose order is a power p^i of the prime number p, if the equation $y^{*p^k}=w^*$ has a solution for every positive integer k, and if x^* is an element of order p^j in G^*, then

$$(Z(G) < G;\ w^*,\ x^*) = w^* \circ x^* = v^{*p^j} \circ x^* \text{ for some } v^* \in G^*$$
$$= v^* \circ x^{*p^j} = 1$$

by (1.1). Now it follows from (1.2) that $w^*=1$. Since G^* is countable, it follows from a theorem of H. Prüfer‡ that G^* is a direct product of cyclic groups.

In Appendix A a uniqueness theorem will be proved in which $G/Z(G)$ does not contain elements of infinite order and need not be a direct product of cyclic groups.

Example 6.5. This is to show that *Theorem* 6.2 *does not hold if $G/Z(G)$ is a countable abelian group but not a direct product of cyclic groups.*

Denote by A a direct product of a countably infinite number of infinite cyclic groups, by $a(1), a(2), \cdots$, a basis of A and by B an abelian group which is generated by elements $b(1), b(2), \cdots$, subject to the relations $b(i+1)^n=b(i)$, for $i=1, 2, \cdots$, and a fixed integer $n>1$. Z is the direct product of A and B.

The abelian group G^* is generated by elements $g(i, j)^*$, for $j=1, 2$ and $i=1, 2, \cdots$, satisfying the relations $g(i+1, j)^{*n}=g(i, j)^*$, for $i=1, 2, \cdots$,

† That these conditions are sufficient is an obvious consequence of well known theorems in the theory of extensions of groups; cf., for example, R. Baer, Mathematische Zeitschrift, vol. 38 (1934), p. 410, formula (3), or p. 391, Theorem 2.

‡ H. Prüfer, Mathematische Zeitschrift, vol. 17 (1923), pp. 48–57.

G^* does not contain elements of finite order not equal to one, and a normal form of the elements in G^* is

$$g(i, 1)^{*h} g(m, 2)^k,$$

with $0 < h < n$, if $1 < i$, and $0 < k < n$, if $1 < m$.

Essentially the only operation $x^* \circ y^*$ of G^* in Z which satisfies the conditions (1) to (4) of Corollary 2.3 is defined by

$$g(i, 1)^{*h} g(m, 2)^{*k} \circ g(i', 1)^{*h'} g(m', 2)^{*k'} = b(i + m' - 1)^{hk'} b(i' + m - 1)^{-h'k}.$$

The group G is generated by adjoining to Z elements $g(i, j)$ which are subject to the relations

$$Z \leqq Z(G), \quad g(i + 1, j)^n = g(i, j), \quad (g(i, 1), g(k, 2)) = b(i + k - 1).$$

The group W is generated by adjoining to Z elements $w(i, j)$ which are subject to the relations

$$Z \leqq Z(W), \quad w(i + 1, j)^n = w(i, j)a(i), \quad (w(i, 1), \quad w(k, 2)) = b(i + k - 1).$$

It follows from (1.2) that both groups G and W have Z as central and G^* as central quotient group, and realize the operation $x^* \circ y^*$. Since G^* does not contain elements other than one of finite order, the conditions of Theorem 6.2 are satisfied with the exception that G^* is not a direct product of cyclic groups. But G and W are not isomorphic. For every class, not equal to Z, of G/Z contains elements which are n^rth powers of elements in G for every positive integer r. But if x is an element in the class $Zw(1, 1)$ of W/Z, then the integers r, such that $y^{n^r} = x$ has a solution y in W, are bounded, as an n^rth power of an element in W which is contained in $Zw(1, 1)$ has the form

$$z^{n^r} a(r)^{n^{r-1}} a(r - 1)^{n^{r-2}} \cdots a(1)w(1, 1), \; z \text{ in } Z.$$

It is a consequence of Theorem 2.1 and Theorem 6.2 that the problem to construct all groups whose centrals are a preassigned abelian group Z and whose central quotient groups are isomorphic to a given direct product G^* of cyclic groups is equivalent to the problem to define all sets of functions

$$x^* \circ y^*, \qquad P(n, x^*)$$

which satisfy the conditions (1) to (4) of Corollary 2.3 and the conditions (2), (3) of Theorem 2.1. If

$$x^* \circ y^*, \quad P(n, x^*) \quad \text{and} \quad x^* \ominus y^*, \quad P'(n, x^*)$$

are two sets of functions of G^* in Z, subject to the mentioned conditions, then they characterize isomorphic groups if, and only if, there exists an automorphism ζ of Z and an automorphism γ of G^* such that

(a) $x^{*\gamma} \circ y^{*\gamma} = (x^* \ominus y^*)^\zeta$ for x^*, y^* in G^*, and

(b) $P(n, x^{*\gamma}) = P'(n, x^*)^\zeta$ for x^* in G_n^*.

That it is not sufficient to assume the existence of one pair of automorphisms γ, ζ satisfying (a) and of another pair of automorphisms γ', ζ', satisfying (b), may be seen from the following example:

Example 6.6. Suppose that Z is a direct product of a cyclic group of order p^3, generated by v, and a cyclic group of order p, generated by u, where p is a prime number not 2, and that G^* is a direct product of two cyclic groups of order p^2 (b^*, c^* a basis of G^*). Admissible sets of functions of G^* in Z may be characterized by the equations

$$b^* \circ c^* = v^p u, \quad P(p^2, b^*) = P(p^2, c^*) = Z^{p^2} v^p u,$$

and another set by the equations

$$b^* \ominus c^* = v^p u, \quad P(p^2, b^*)' = P(p^2, c^*)' = Z^{p^2} v^p u^{-1}.$$

Clearly there exists a pair of automorphisms γ, ζ which satisfies (a) and another pair which satisfies (b) but none which satisfies both (a) and (b).

7. **Types of subgroups.** Suppose that G is a group with abelian central quotient group, and that the subgroups S and T of G are situated between the central and the commutator group of G. Assume furthermore that G/S is a direct product of cyclic groups. Then it follows from Theorem 5.1 that there exists an automorphism of G which maps S upon T (that is, that S and T are *isotype* in G) if, and only if, there exists an isomorphism σ of S upon T and an isomorphism λ of G/S upon G/T such that

(a) $(S<G; x^*, y^*)^\sigma = (T<G; x^{*\lambda}, y^{*\lambda})$ for x^*, y^* in G/S and

(b) $P(S<G; n, x^*)^\sigma = P(T<G; n, x^{*\lambda})$ for x^* in $(G/S)_n$.

If, in particular, G is an abelian group, then condition (a) may be omitted, and this characterization of the types of subgroups applies to any subgroup whose quotient group is a direct product of cyclic groups. It applies therefore to all subgroups, if, for example, G is a group with a finite number of generators or if the orders of the elements in G are bounded.

8. **Conformality.** Two groups G and H are said to be conformal, if they contain for every positive integer n the same number of elements of order n and also the same number of elements of infinite order. Since (as obvious examples show) infinite abelian groups without elements of infinite order may be conformal without being isomorphic, this definition of conformality is too wide for our purposes, and therefore the following definition may be adopted.

DEFINITION 8.1. *The groups G and H with abelian central quotient groups are said to be conformal if there exist groups S and T between the central and the commutator groups of G and H, respectively, and an S-T-transformation of G*

into H which is a one-one correspondence between G and H and satisfies condition (6) *of Theorem* 4.1. *Such an S-T-transformation may be termed an S-T-conformality of G upon H.*

Since S-T-conformalities induce isomorphisms in S and G/S and satisfy condition (5′) of §4, conformal groups are also conformal in the wider sense mentioned above. It is a consequence of condition (6) of Theorem 4.1 that conformalities between abelian groups are isomorphisms.

Conformality is a symmetric and reflexive relation. But in general it is not transitive as follows from Example 8.4.

Suppose now that $C(G) \leqq S \leqq Z(G)$, $C(H) \leqq T \leqq Z(H)$ and that G/S is a direct product of cyclic groups. Then it follows from the results of §4, that there exists an S-T-conformality of G upon H if, and only if, there exists an isomorphism σ of S upon T and an isomorphism λ of G/S upon H/T such that

(a) $P(S<G; n, x^*)^\sigma = P(T<H; n, x^{*\lambda})$ for x^* in $(G/S)_n$ and

(b) $(S<G; x^*, y^*)^{2^{m-1}\sigma} = (T<H; x^{*\lambda}, y^{*\lambda})^{2^{m-1}}$ for x^*, y^* in $(G/S)_{2^m}$.

Exactly those pairs σ, λ which satisfy the conditions (a) and (b) may be induced by S-T-conformalities of G upon H.

As condition (b) is not very restrictive, this shows that conformality depends essentially on the P-functions.

THEOREM 8.2. *If G is a group with abelian central quotient group, and if $G/C(G)$ is a direct product of cyclic groups, then the following three properties of G are equivalent*:

(a) *G is conformal to an abelian group.*

(b) *If x and y are elements of G and if n is a positive integer such that x^n and y^n are elements of $C(G)$, then $x^n y^n = (xy)^n$.*

(c) $(C(G)<G; x^*, y^*)^{2^{m-1}} = 1$ *for* x^*, y^* *in* $(G/C(G))_{2^m}$.

Proof. If (a) holds, then there exists a group S between $C(G)$ and $Z(G)$, an abelian group H and a subgroup T of H, and an S-T-conformality ϕ of G upon H. If n is a positive integer, and if x and y are elements in G such that x^n and y^n are elements in $C(G)$, then x^n and y^n are elements in S; hence

$$\begin{aligned}((xy)^n)^\phi &= (x^\phi y^\phi)^n = (x^\phi)^n (y^\phi)^n \\ &= (x^n)^\phi (y^n)^\phi = (x^n y^n)^\phi,\end{aligned}$$

since $G^\phi = H$ is an abelian group, condition (5′) holds, and ϕ is a homomorphism on S. Now (b) is a consequence of the fact that ϕ defines a one-one correspondence between G and H.

If (b) holds true, then it follows from (1.3) that for any two elements x and y such that $C(G)x$ and $C(G)y$ are elements of $(G/C(G))_{2^m}$ the following equalities hold:

$$x^{2^m}y^{2^m} = (xy)^{2^m} = (C(G) < G; C(G)x, C(G)y)^{2^{m-1}}x^{2^m}y^{2^m};$$

and (c) is therefore a consequence of (b).

Suppose now that (c) is satisfied. Then it follows from Theorem 2.1 that there exists an abelian group H, which contains $C(G) = S$ as a subgroup, and an isomorphism γ of $G/C(G)$ upon H/S such that $P(S<H; n, x^{*\gamma}) = P(C(G)<G; n, x^*)$ for x^* in $(G/C(G))_n$. It follows, from the fact mentioned in the beginning of this section, that there exists a $C(G)$-S-conformality of G upon H which maps every element in $C(G) = S$ upon itself and induces γ in $G/C(G)$. Hence (a) is a consequence of (c).

Note that the existence of a basis of $G/C(G)$ has been needed only to prove that (a) is a consequence of (c).

THEOREM 8.3. *If G is a group with abelian central quotient group, if $G/C(G)$ is a direct product of cyclic groups, and if H and H' are abelian groups which are both conformal to G, then H and H' are isomorphic.*

Proof. From the assumptions there exist subgroups S and S' between $C(G)$ and $Z(G)$, an S-S^ϕ-conformality ϕ of G upon H, and an S'-$S'^{\phi'}$-conformality ϕ' of G upon H'. Since H and H' are abelian groups, ϕ and ϕ' are isomorphisms in $C(G)$, it follows from condition (6) of Theorem 4.1 that ϕ is a $C(G)$-$C(G)^\phi$-conformality of G upon H, and ϕ' is a $C(G)$-$C(G)^{\phi'}$-conformality of G upon H'. Consequently there exists an isomorphism σ of $C(G)^\phi = T$ upon $C(G)^{\phi'} = T'$ and an isomorphism λ of H/T upon H'/T' such that

$$P(T < H; n, x^*)^\sigma = P(T' < H'; n, x^{*\lambda})$$

for x^* in $(H/T)_n$. Since $G/C(G)$, H/T, and H'/T' are isomorphic groups and therefore direct products of cyclic groups, and since H and H' are abelian groups, it follows from Theorem 5.1 that there exists an isomorphism of H upon H' which induces σ in T and λ in H/T.

This proof together with the considerations in §7 makes it clear that the investigation of the functions $P(n, x^*)$ is a problem which is equivalent to the problem of characterization of the classes of isotype subgroups of an abelian group.

Example 8.4. This is to show that *the Theorem 8.3 is no longer true, when the condition of the existence of a basis of $G/C(G)$ is omitted*, and this proves incidentally that *the conformality relation is not transitive.*

Let A be the direct product of infinite cyclic groups with the basis $a(1)$, $a(2), \cdots$, and let B be the abelian group generated by the elements $b(1)$, $b(2), \cdots$ subject to the relations

$$b(i+1)^n = b(i),$$

where $1<n$ is a fixed odd integer. Z denotes the direct product of A and B.

A^* denotes the direct product of infinite cyclic groups with the basis $a^*(1), a^*(2), \cdots$ and B^* the abelian group generated by elements $b^*(i, j)$, for $i=1, 2, \cdots$ and $j=1, 2$, subject to the relations

$$b^*(i+1, j)^n = b^*(i, j).$$

G^* denotes the direct product of A^* and B^*.

An operation $x^* \circ y^*$ of G^* in Z, satisfying the conditions (1) to (4) of Corollary 2.3, is characterized by the equations

$$\begin{aligned} a^*(2i-1) \circ a^*(2i) &= a(i), \\ a^*(i) \circ a^*(j) &= 1, \quad \text{for} \quad i<j, \qquad (i, j) \neq (2k-1, 2k), \\ b^*(i, j) \circ a^*(k) &= 1, \\ b^*(i, 1) \circ b^*(k, 2) &= b(i+k-1). \end{aligned}$$

The group G may be the group, generated by adjoining to Z elements $u(1), u(2), \cdots, v(i, j)$, for $i=1, 2, \cdots; j=1, 2$, subject to the relations

$$\begin{aligned} Z \leqq Z(G), \qquad (u(i), u(j)) &= a^*(i) \circ a^*(j), \qquad (u(i), v(k, j)) = 1, \\ (v(i, 1), v(k, 2)) &= b^*(i, 1) \circ b^*(k, 2), \\ v(i+1, j)^n &= v(i, j). \end{aligned}$$

G is an extension of Z by G^* which realizes the operation $x^* \circ y^*$. Hence $Z=C(G)=Z(G)$.

It is easily seen that G is conformal to the abelian group H which is the direct product of Z and G^*.

Denote by U' the direct product of infinite cyclic groups generated by the elements $u'(1), u'(2), \cdots$, by B' the direct product of the infinite cyclic groups, generated by the elements $b'(i, j)$, for $i=1, 2, \cdots, j=1, 2$, and by A' the subgroup of B' which is generated by the elements $a'(2(i-1)+j) = b'(i, j)b'(i+1, j)^{-n}$ and is the direct product of the cyclic groups generated by the elements $a'(k)$. Let H' be the direct product of U', B', and B, and denote by Z' the direct product of A' and B.

A Z-Z'-conformality ϕ of G upon H' is defined by

$$\begin{aligned} u'(i) &= u(i)^\phi, \qquad b'(i, j) = v(i, j)^\phi, \\ a'(i) &= a(i)^\phi, \qquad b(i) = b(i)^\phi, \end{aligned}$$

since ϕ defines an isomorphism of Z upon Z' and of G^* upon H'/Z'. But H and H' are abelian groups which are not isomorphic,† since H is a direct prod-

† Cf. R. Baer, Duke Mathematical Journal, vol. 3 (1937), pp. 69–122, Corollary 2.9.

uct of cyclic groups by three groups of the type of B, whereas H' is a direct product of cyclic groups by one group of the type of B.

It is a special case of Theorem 8.2 that the group G is conformal to an abelian group, if $C(G) \leqq Z(G)$, and if $G/C(G)$ is a direct product of cyclic groups and does not contain elements of even order. Suppose on the other hand that $C(G) \leqq S \leqq Z(G)$ and that 2^m with $0 < m$ is the l.c.m. of the orders of the elements in G/S. If G is S-$\cdots$-conformal to an abelian group, and if w^* is an element of order 2^m in G/S, then

$$1 = (S < G;\ w^*,\ x^*)^{2^{m-1}} = (S < G;\ w^{*2^{m-1}},\ x^*)$$

for every x^* in G/S; and it follows from (1.1) that the elements in $w^{*2^{m-1}}$ are elements of $Z(G)$. Since $1 \neq w^{*2^{m-1}}$, this implies that $S \neq Z(G)$.

It is fairly obvious how to construct groups of order a power of 2 which are conformal to an abelian group. Let, for example, n be an integer, $(1 < n)$, S a cyclic group of order 2^{n-1} generated by s, and G^* a direct product of two cyclic groups of order 2^n, (u^*, v^* a basis of G^*). The group G with $C(G) \leqq S \leqq Z(G)$ which realizes the functions characterized by

$$u^* \circ v^* = s, \qquad P(2^n, u^*) = P(2^n, v^*) = 1$$

is, by Theorem 8.2, conformal to an abelian group, though not itself abelian. If in this construction it had been assumed that S was of order 2^n, then the group G would not be conformal to any abelian group, since $S = C(G) = Z(G)$, and it may be computed that the numbers of the elements of a given order in G are not the same as the corresponding numbers in any abelian group.

9. **Groups with isomorphic commutator groups and isomorphic central quotients groups.** If ϕ is a homomorphism of the group G upon the whole group H, then $C(G)^\phi = C(H)$ and $Z(G)^\phi \leqq Z(H)$. If S is a normal subgroup of G, then S^ϕ is a normal subgroup of H, and ϕ induces an isomorphism of G/S upon H/S^ϕ, if S is the complete origin of S^ϕ.

LEMMA 9.1. *If G is a group with abelian central quotient group, and if ϕ is a homomorphism of G upon the (whole) group H which induces an isomorphism in $C(G)$, then $Z(G)^\phi = Z(H)$, $Z(G)$ is the complete origin of $Z(H)$, and ϕ induces an isomorphism in $G/Z(G)$.*

Proof. Let w be any element in G such that w^ϕ is an element in $Z(H)$. If x is any element in G, then $(w, x)^\phi = (w^\phi, x^\phi) = 1$, consequently $(w, x) = 1$ for every x in G; that is, w is an element in $Z(G)$. This proves the lemma.

That the converse does not hold, may be seen from the following example.

Let p be a prime number not 2, and let Z and G^* each be a direct product of three cyclic groups of order p. If z, z', and z'' form a basis of Z, and u^*, v^*,

and w^* a basis of G^*, then an operation $x^* \circ y^*$ of G^* in Z which satisfies the conditions (1) to (4) of Corollary 2.3 is characterized by the equations

$$u^* \circ v^* = z, \qquad v^* \circ w^* = z', \qquad w^* \circ u^* = z''.$$

The functions $P(p, x^*) = 1$, for x^* in G^*, satisfy the conditions of Theorem 2.1 with the operation $x^* \circ y^*$. It is therefore a consequence of Theorem 2.1, (1.2), and Theorem 6.1 that there exists one, and essentially only one, group G such that $Z = Z(G)$ (or $C(G)$) and $G/Z = G^*$ which realizes the functions $x^* \circ y^*$ and $P(p, x^*)$.

Denote by M the subgroup of G generated by z, and put $H = G/M$. Then it follows from (1.2) that $Z(H) = C(H) = Z(G)/M$, but this homomorphism of G upon H is not an isomorphism in $C(G)$.

THEOREM 9.2. *Suppose that G and G' are groups whose central quotient groups are abelian. Then there exists an isomorphism ψ of $C(G)$ upon $C(G')$ and an isomorphism λ of $G/Z(G)$ upon $G'/Z(G')$ with the property that*

$$(Z(G) < G;\ x^*,\ y^*)^{\psi} = (Z(G') < G';\ x^{*\lambda},\ y^{*\lambda})$$

for x^*, y^* in $G/Z(G)$ if, and only if, there exists a group H with abelian central quotient group and homomorphisms ϕ and ϕ' of H upon G and G', respectively, which induce isomorphisms in $C(H)$.

Proof. Suppose first that there exists a group H with abelian central quotient group and homomorphisms ϕ and ϕ' of H upon G and G', respectively, which induce isomorphism in $C(H)$. Then it follows from Lemma 9.1 that $Z(H)^{\phi} = Z(G)$, $Z(H)^{\phi'} = Z(G')$, and that ϕ and ϕ' induce isomorphisms in $H/Z(H)$. Thus ϕ^{-1} is an isomorphism in $C(G)$ and in $G/Z(G)$, and $\psi = \phi^{-1}\phi'$ is an isomorphism of $C(G)$ upon $C(G')$, $\lambda = \phi^{-1}\phi'$ is an isomorphism of $G/Z(G)$ upon $G'/Z(G')$.

If x and y are two elements in G, then there exist some elements u and v in H such that $u^{\phi} = x$, $v^{\phi} = y$. Consequently

$$(x, y)^{\psi} = (u, v)^{\phi'} = (u^{\phi'}, v^{\phi'}),$$
$$(Z(G)x, Z(G)y)^{\psi} = ((Z(G)x)^{\lambda}, (Z(G)y)^{\lambda}),$$

and the condition is a sufficient one.

Suppose now that there exists an isomorphism ψ of $C(G)$ upon $C(G')$ and an isomorphism λ of $G/Z(G)$ upon $G'/Z(G')$ such that

(i) $(Z(G) < G;\ x^*,\ y^*)^{\psi} = (Z(G') < G';\ x^{*\lambda} y^{*\lambda})$ for x^*, y^* in $G/Z(G)$.

Since $G/C(G)$ is an abelian group, there exist a direct product G^* of infinite cyclic groups and a homomorphism γ of G^* upon $G/C(G)$. There exist by

Theorem 2.1 and by Theorem 6.1 one, and essentially only one, group K, such that $C(K)=C(G)\leqq Z(K)$, and an isomorphism κ of $K/C(K)$ upon G^* such that

$$(C(K) < K;\ x^*,\ y^*) = (C(G) < G;\ x^{*\kappa\gamma},\ y^{*\kappa\gamma})$$

for x^*, y^* in $K/C(K)$. Since $K/C(K)$ is a direct product of infinite cyclic groups, $Z(K)/C(K)$ is also a direct product of infinite cyclic groups,† consequently $Z(K)$ is the direct product of $C(K)$ and a group D which is a direct product of infinite cyclic groups. Since $\kappa\gamma$ is a homomorphism of $K/C(K)$ upon $G/C(G)$, since $K/C(K)$ is a direct product of infinite cyclic groups, it follows from the above equation and Theorem 5.1 that there exists a homomorphism ρ of K upon G which leaves the elements in $C(K)=C(G)$ invariant and induces $\kappa\gamma$ in $K/C(K)$.

Similarly there exists a group K', such that $C(G')=C(K')\leqq Z(K')$, $K'/C(K')$ is a direct product of infinite cyclic groups, and $Z(K')$ is the direct product of $C(K')$ and a group D' which is a direct product of infinite cyclic groups; and there exists a homomorphism ρ' of K' upon G' which leaves all the elements in $C(K')=C(G')$ invariant.

Since $G/Z(G)$ is an abelian group, there exists a direct product H^* of infinite cyclic groups and a homomorphism ι of H^* upon $G/Z(G)$. Denote by S the direct product of D, D', and $C(G)$, and denote by H the essentially uniquely determined group such that $C(H)\leqq S\leqq Z(H)$ and such that there exists an isomorphism τ of H/S upon H^* which satisfies the equation

$$(S < H;\ x^*,\ y^*) = (Z(G) < G;\ x^{*\tau\iota},\ y^{*\tau\iota})$$

for x^*, y^* in H/S.

Since H/S is a direct product of infinite cyclic groups, $H/C(H)$ is the direct product of $S/C(H)$ and a direct product D^* of infinite cyclic groups.

By Lemma 9.1, ρ induces an isomorphism of $K/Z(K)$ upon $G/Z(G)$; consequently the isomorphism ρ^{-1} is well defined on $G/Z(G)$. Since D^* is a direct product of infinite cyclic groups and since D^* represents exactly H/S, there exists a homomorphism σ of $H/C(H)$ upon $K/C(K)$ with the following properties: $x^{*\sigma}=1$, if $x^*=C(H)d'$ and d' in D'; $x^{*\sigma}=C(K)d$, if $x^*=C(H)d$ and d is an element of D; and $Z(K)^{*\sigma}=(Sx^*)^{\tau\iota\rho^{-1}}$, if x^* is an element of D^*.

Since ρ induces $\kappa\gamma$ in $K/C(K)$, it follows that

$$(S < H;\ x^*,\ y^*) = (Z(K) < K;\ x^{*\tau\iota\rho^{-1}},\ y^{*\tau\iota\rho^{-1}})$$

for x^*, y^* in H/S. The homomorphism σ therefore satisfies

† R. Baer, Duke Mathematical Journal, vol. 3 (1937), pp. 69–122, especially the remark to Corollary 8.9.

$$(C(K) < K;\ x^{*\sigma},\ y^{*\sigma}) = (C(H) < H;\ x^*,\ y^*)$$

for x^*, y^* in $H/C(H)$.

By Lemma 9.1, ρ' induces an isomorphism of $K'/Z(K')$ upon $G'/Z(G')$; consequently the isomorphism ρ'^{-1} is well defined on $G'/Z(G')$. There exists, therefore, a homomorphism σ' of $H/C(H)$ upon $K'/C(K')$ with the following properties: $x^{*\sigma'} = C(K')d'$, if $x^* = C(H)d'$ and d' is an element in D'; $x^{*\sigma'} = 1$, if $x^* = C(H)d$ and d is an element in D; and $Z(K')x^{*\sigma'} = (Sx^*)^{\tau_i\lambda\rho'^{-1}}$, if x^* is an element of D^*.

Since ρ' is a homomorphism of K' upon G' which leaves all the commutators invariant, it follows from condition (i) that

$$\begin{aligned}(C(K') < K';\ x^{*\sigma'},\ y^{*\sigma'}) &= (Z(K') < K';\ Z(K')x^{*\sigma'},\ Z(K')y^{*\sigma'})\\ &= (Z(K') < K';\ (Sx^*)^{\tau_i\lambda\rho'^{-1}},\ (Sy^*)^{\tau_i\lambda\rho'^{-1}})\\ &= (Z(G') < G';\ (Sx^*)^{\tau_i\lambda},\ (Sy^*)^{\tau_i\lambda})\\ &= (Z(G) < G;\ (Sx^*)^{\tau_i},\ (Sy^*)^{\tau_i})^{\psi}\\ &= (S < H;\ Sx^*, Sy^*)^{\psi} = (C(H) < H;\ x^*,\ y^*)^{\psi}\end{aligned}$$

for every x^*, y^* in $H/C(H)$.

Since $H/C(H)$ is a direct product of infinite cyclic groups, it follows from Theorem 5.1 that there exists a homomorphism β of H upon K which induces σ in $H/C(H)$ and leaves the elements in $C(H) = C(K)$ invariant, and that there exists a homomorphism β' of H upon K' which induces σ' in $H/C(H)$ and ψ in $C(H) = C(G)$. $\beta\rho$ and $\beta'\rho'$ are homomorphisms of H upon G and G', respectively. $\beta\rho$ leaves the elements in $C(H) = C(K) = C(G)$ invariant, and $\beta'\rho'$ induces in $C(H)$ the isomorphism ψ. This completes the proof.

In fact slightly more has been proved, namely, that the group H, which meets the requirements of the Theorem 9.2, may always be chosen in such a way that $H/C(H)$ is a direct product of infinite cyclic groups.

The proof of this theorem shows the advantage of treating finite and infinite groups at the same time. For even if the groups G and G' of the theorem are finite, it may prove difficult to construct a group H which meets all the requirements without recourse to infinite groups.

10. **Automorphisms.** If G is a group with abelian central quotient group and S a subgroup of G such that $C(G) \leqq S \leqq Z(G)$, then denote by $\Theta(S<G)$ the group of all those (proper) automorphisms of G which map S upon itself. If S is, for example, the central or the commutator group of G, then $\Theta(S<G)$ is the group of all automorphisms of G.

Those automorphisms of G which leave invariant every element in S and every element in G/S form a normal subgroup $\Omega(S<G)$ of $\Theta(S<G)$. If ϕ is

any automorphism in $\Omega(S<G)$, x an element in G and s an element in S, then

$$x^{\phi}x^{-1} = (sx)^{\phi}(sx)^{-1} = (Sx)^{\phi}(Sx)^{-1} = (\phi, Sx)$$

is an element in S; and if x and y are two elements in G, then

$$\begin{aligned}(\phi, Sxy) &= (xy)^{\phi}(xy)^{-1} = x^{\phi}y^{\phi}y^{-1}x^{-1}\\ &= x^{\phi}(\phi, Sy)x^{-1} = (\phi, Sx)(\phi, Sy),\end{aligned}$$

and conversely every homomorphism of G/S into S may be realized this way. Since finally

$$\begin{aligned}(\phi\gamma, Sx) &= x^{\phi\gamma}x^{-1} = x^{\phi\gamma}(x^{-1})^{\gamma}x^{\gamma}x^{-1}\\ &= (\phi, Sx)^{\gamma}(\gamma, Sx) = (\phi, Sx)(\gamma, Sx),\end{aligned}$$

as γ belongs to $\Omega(S<G)$ and (ϕ, Sx) to S, it follows that $\Omega(S<G)$ *is essentially equal to the group of all homomorphisms of* G/S *into* S.

Two automorphisms in $\Theta(S<G)$ belong to the same class of $\Lambda(S<G) = \Theta(S<G)/\Omega(S<G)$ if, and only if, they induce the same automorphisms in S and in G/S. Thus $\Lambda(S<G)$ is essentially equal to a subgroup of the group $\Pi(S, G/S)$ of all the pairs (σ, λ) of automorphisms σ of S and λ of G/S. If the pair (σ, λ) in $\Pi(S, G/S)$ is realized by an automorphism in $\Theta(S<G)$, then it follows that

(a) $(S<G; x^*, y^*)^{\sigma} = (S<G; x^{*\lambda}, y^{*\lambda})$ for x^*, y^*, in G/S, and

(b) $P(S<G; n, x^*)^{\sigma} = P(S<G; n, x^{*\lambda})$ for x^* in $(G/S)_n$;

and these conditions are sufficient for the realizability, provided G/S is a direct product of cyclic groups, as follows from Theorem 5.1.

Suppose now that ϕ is an automorphism in $\Omega(S<G)$ and that the automorphism γ of G induces in S the automorphism σ and in G/S the automorphism λ. Then

$$\begin{aligned}(\gamma^{-1}\phi\gamma, Sx) &= x^{\gamma^{-1}\phi\gamma}x^{-1} = (x^{\gamma^{-1}\phi}(x^{\gamma^{-1}})^{-1})^{\gamma}\\ &= (\phi, Sx^{\gamma^{-1}})^{\gamma} = (\phi, (Sx)^{\lambda^{-1}})^{\sigma};\end{aligned}$$

consequently it follows that the class of $\Lambda(S<G)$ which is characterized by the pair (σ, λ) in $\Pi(S, G/S)$ induces, in the abelian group $\Omega(S<G)$, the automorphism defined by

$$(\phi^{(\sigma,\lambda)}, x^*) = (\phi, x^{*\lambda^{-1}})^{\sigma}$$

for x^* in G/S.

In order to characterize the extension $\Theta(S<G)$ of $\Omega(S<G)$ by $\Lambda(S<G)$ which realizes the above mentioned automorphisms in $\Omega(S<G)$, it is necessary to compute the so-called factor sets. A method for their calculation may be indicated for groups G, S which satisfy the conditions:

(i) $C(G) \leqq S \leqq Z(G)$.

(ii) $G/S = G^*$ is a direct product of cyclic groups.

(iii) If x and y are elements of G and n is a positive integer such that x^n and y^n are elements of S, then $x^n y^n = (xy)^n$.

First we define in S for every positive integer n a function $s^{n^{-1}}$ as follows: $1 = 1^{n^{-1}}$; if the equation $x^n = s$ has a solution in S, then $s^{n^{-1}}$ is one of these solutions x; if s is not contained in S^n, then $s^{n^{-1}}$ is not defined.

Second an ordered set B is chosen in G which represents exactly a basis B^* of G/S, and the functions

$$s(x), \quad n(x, b), \quad a(x, y) = a(Sx, Sy), \quad c(x, y) = c(Sx, Sy),$$

for x, y in G and b in B, introduced in §3, are defined with reference to the ordered set B.

In defining another multiplication of the elements of G by the equation

$$x^* y = c(x, y)^{-1} xy,$$

for x, y in G, G is transformed into an abelian group H, and S is a subgroup of H, since $c(x, y) = 1$ for x, y in S. Thus the identical mapping of the elements in G is an S-S-conformality of G upon H. Furthermore $G/S = H/S$ and $P(S<G; n, x^*) = P(S<H; n, x^*)$ for x^* in $G_n{}^*$. Finally $\Omega(S<G) = \Omega(S<H)$ and $\Lambda(S<G) \leqq \Lambda(S<H)$, as follows from the conditions (a) and (b) mentioned above.

Suppose now that (σ, λ) is a pair in $\Pi(S, G/S)$ which may be realized by some automorphism in $\Theta(S<G)$. Then

$$b^{*\lambda} = \prod_{d^* \in B^*} d^{*n(b^*, d^*, \lambda)},$$

where almost every $n(b^*, d^*, \lambda) = 0$ and $0 \leqq n(b^*, d^*, \lambda) < n(d^*)$, if d^* is of finite order $n(d^*)$. If b^* is an element of finite order $n = n(b^*)$, and if d in B always represents the element d^* in B^*, then it follows from the realizability of (σ, λ) that the elements

$$(b^n)^\sigma \prod_{d \in B} d^{-n n(b^*, d^*, \lambda)} = s(b, \sigma, \lambda)$$

are nth powers of elements in S. Consequently, an automorphism $\gamma = \gamma(\sigma, \lambda)$ in $\Theta(S<G)$ is defined by the equations

$$x^\gamma = x^\sigma \text{ for } x \text{ in } S,$$

$$b^\gamma = t(b, \sigma, \lambda) \prod_{d \in B} d^{n(b^*, d^*, \lambda)},$$

where $t(b, \sigma, \lambda) = 1$, if b^* is of infinite order, and $t(b, \sigma, \lambda) = s(b, \sigma, \lambda)^{n(b^*)^{-1}}$, if b^* is of the finite order $n(b^*)$.

It is easily seen that there exists an automorphism in $\Theta(S<H)$ which has the same values as γ on S and on B. But these automorphisms do not constitute the same mapping of the elements which form both groups G and H.

Thus a representative of every class of $\Lambda(S<G)$ has been chosen, and the factor sets are expressions

$$\gamma(\sigma, \lambda)\gamma(\sigma', \lambda')\gamma(\sigma\sigma', \lambda\lambda')^{-1} = a(\sigma, \lambda; \sigma', \lambda')c(\sigma, \lambda; \sigma', \lambda')$$

where $a(\cdots)$ is the factor set, calculated for the group H, and $c(\cdots)$ is an expression in the $(S<G; b^*, d^*)$.

If all the elements of $G/Z(G)$ are of finite odd order, then the investigations of Appendix B will produce better results. But the methods of Appendix B break down in some of those cases in which the method of this section may be applied.

Appendix A. A uniqueness theorem for p-adic groups. It is the object of this appendix to prove a uniqueness theorem which is not contained in Theorem 6.2. Because of Example 6.5 only groups without elements of infinite order will be considered, and by Remark 6.4 it will be necessary to consider groups whose central quotient group is infinite but not countable.

Groups whose central quotient group is abelian are direct products of p-groups (that is, groups which contain only elements of order a power of the prime number p) if, and only if, they do not contain elements of infinite order.† Consequently only p-groups will be considered in this appendix.

The p-group G will be called *p-adic* if the cross cut of its subgroups G^{p^i}, for $i = 1, 2, \cdots$, consists of the group unit element only. The p-group G is said to be *dense in the p-group H*, if $G \leqq H$ and every class of H/H^p contains elements of G. Finally G is termed a *closed p-adic group* if G is p-adic and if G is the only p-adic group in which G is dense.

LEMMA A.1. (a) *Every p-adic group is dense in one and essentially only one closed p-adic group.*

(b) *If the p-adic group G is dense in the closed p-adic group $\overline{G}$, and if the p-adic group H is dense in the closed p-adic group $\overline{H}$, then every isomorphism γ of G upon the whole group H is induced by one and only one isomorphism of $\overline{G}$ upon $\overline{H}$.*

---

† Cf., for example, R. Baer, Compositio Mathematica, vol. 1 (1934), pp. 254–283, especially the lemma on p. 261.

Similar statements have been proved before by similar methods,† and it will therefore be sufficient to give the following indication of a proof of this lemma. If G is a p-adic group, then consider the set $S(G)$ of all the sequences S_i with the following properties:

(a) S_i is a class of G/G^{p^i}.

(b) $S_{i+1} \leqq S_i$.

(c) The orders of the elements S_i of the groups G/G^{p^i} are bounded. (The upper bounds for the orders may be different for different sequences.)

If the multiplication in $S(G)$ is defined in the obvious way, $S(G)$ becomes a closed p-adic group. In mapping the element x of G upon the sequence $G^{p^i}x$ an isomorphism of G upon a dense subgroup of $S(G)$ is defined. Now it is fairly obvious how to work out a proof of the lemma.

If G is a p-adic group, then the essentially uniquely determined closed p-adic group $\overline{G}$, in which G is dense, may be called the *p-adic closure of G*.

The following example shows the existence of closed p-adic groups whose central quotient group is abelian but not a direct product of cyclic groups.

Denote by Z_i and Z_i^* cyclic groups of order p^i, generated by z_i and z_i^*, respectively; and let Z be the direct product of the groups Z_i, for $i=1, 2, \cdots$, and Z^* the direct product of the groups Z_i^*, for $i=1, 2, \cdots$. An operation $x^* \circ y^*$ and functions $P(p^i, x^*)$ of Z^* in Z, which satisfy the conditions (1) to (4) of Corollary 2.3 and the conditions (1) to (3) of Theorem 2.1, may be defined by the following equations:

$$z_i^* \circ z_{i+1}^* = z_i, \quad z_i^* \circ z_{i+h}^* = 1 \quad (\text{for } 1 < h); \quad P(p^i, z_i^*) = 1.$$

Thus there exists one, and essentially only one, group G which satisfies $Z = C(G) = Z(G)$, $Z^* = G/Z$, and which realizes the functions $x^* \circ y^*$, $P(p^i, x^*)$. G is clearly a p-adic group.

If $\overline{G}$ is the p-adic closure of G, then $Z(\overline{G})$ is the p-adic closure of $Z(G)$, and $\overline{G}/Z(\overline{G})$ is the p-adic closure of $G/Z(G)$. It is well known that these abelian groups are not direct products of cyclic groups.

THEOREM A.2. *Suppose that G is a closed p-adic group with abelian central quotient group. Then G and H are isomorphic groups if, and only if,*

(1) *H is a closed p-adic group;*

(2) *there exists an isomorphism ζ of $Z(G)$ upon $Z(H)$ and an isomorphism λ of $G/Z(G)$ upon $H/Z(H)$ which satisfy the conditions:*

(a) $(Z(G) < G; x^*, y^*)^\zeta = (Z(H) < H; x^{*\lambda}, y^{*\lambda})$ *for x^*, y^* in $G/Z(G)$;*

(b) $P(Z(G) < G; p^i, x^*)^\zeta = P(Z(H) < H; p^i, x^{*\lambda})$ *for x^* in $(G/Z(G))_{p^i}$.*

† R. Baer, Journal für die reine und angewandte Mathematik, vol. 160 (1929), pp. 208–226. H. Freudenthal, Compositio Mathematica, vol. 4 (1937), pp. 145–234.

If condition (1) *is satisfied by* H, *then exactly those pairs of isomorphisms* ζ, λ *which satisfy* (a) *and* (b) *may be induced by isomorphisms of* G *upon* H.

Proof. The necessity of the conditions appearing in either of the statements of the theorem is fairly obvious. Suppose now that condition (1) is satisfied by H, and that ζ, λ is a pair of isomorphisms satisfying the conditions (a) and (b). It shall be proved that there exists an isomorphism of G upon H which induces ζ in $Z(G)$ and λ in $G/Z(G)$.

Since $G^* = G/Z(G)$ is an abelian p-group, there exists a direct product D^* of cyclic groups which is dense in G^*.† Then $D^{*\lambda}$ is dense in the abelian p-group $G^{*\lambda} = H/Z(H)$. If D is the group such that $Z(G) \leqq D \leqq G$, $D/Z(G) = D^*$, and if D' is the group such that $Z(H) \leqq D' \leqq H$, and $D'/Z(H) = D^{*\lambda}$, then D is dense in G and D' is dense in H. It is now a consequence of Theorem 5.1 and the conditions (a), (b) that there exists an isomorphism δ of D upon D' which induces ζ in $Z(G)$ and λ in D^*; and it follows from Lemma A.1, (b) that there exists a uniquely determined isomorphism γ of G upon the closed p-adic group H which induces δ in D. Clearly γ induces ζ in $Z(G)$.

Applying the argument used in Remark 6.4 we can show that $G/Z(G)$ is a p-adic group. Consequently there exists at most one isomorphism of $G/Z(G)$ upon $H/Z(H)$ which induces a given isomorphism in the group D^*, dense in G^*. Since γ and λ are equal isomorphisms on D^*, this implies, therefore, that γ induces λ in the whole group $G/Z(G)$.

It may be noted that by the method applied in the proof even the following slightly more general statements may be proved:

COROLLARY A.3. *Suppose that* G *and* H *are closed* p*-adic groups and that* $C(G) \leqq S \leqq Z(G)$, $C(H) \leqq T \leqq Z(H)$. *Then there exists an isomorphism of* G *upon* H *which maps* S *upon* T *if, and only if, there exists an isomorphism* σ *of* S *upon* T *and an isomorphism* λ *of* G/S *upon* H/T *which satisfy the conditions* (a) *and* (b) *of Theorem* 5.1. *If furthermore* G/S *is a* p*-adic group, then exactly the pairs* σ, λ, *satisfying* (a) *and* (b) *of Theorem* 5.1 *may be realized by isomorphisms of* G *upon* H *which map* S *upon* T.

Appendix B. Conformalities which preserve the automorphisms. We prove the following theorem:

THEOREM B.1. *If* G *is a group with abelian central quotient group, if the subgroup* S *of* G *does not contain elements of order* 2, *and if* $C(G) \leqq S^2 \leqq S \leqq Z(G)$, *then there exists an abelian group* H *and a one-one correspondence* ϕ *which maps* G *upon the whole group* H *and satisfies the conditions*:

(1) $(xy)^\phi = x^\phi y^\phi$ *for* x *in* $Z(G)$ *and* y *in* G.

† R. Baer, American Journal of Mathematics, vol. 59 (1937), pp. 99–117, §1.

(2) $(x^n)^\phi = (x^\phi)^n$ *for* x *in* G *and any integer* n.

(3) ϕ *induces a homomorphism of* G *upon* H/S^ϕ.

(4) *If* γ *is a (proper or improper) automorphism of* G *such that* $S^\gamma \leqq S$, *then* $\phi^{-1}\gamma\phi$ *is an automorphism of* H.

Note that ϕ induces isomorphisms in $Z(G)$ and in G/S.

Proof.† Since $C(G) \leqq S^2$, and since S does not contain elements of order 2, there exists to every pair x, y of elements in G a uniquely determined element $f(x, y)$ in S such that $f(x, y)^2 = (x, y)$. Since $f(x, y)$ is the only solution of this equation in S, and since S is an abelian group, it follows that

$$f(x, x) = f(x, y)f(y, x) = 1, \qquad f(x, yz) = f(x, y)f(x, z);$$

and

$$f(x, y) = 1$$

if, and only if, $(x, y) = 1$.

Denote now by ϕ a one-one correspondence which maps G upon a set H of elements. In H a multiplication may be defined by the equation

$$x^\phi y^\phi = (f(y, x)xy)^\phi$$

for x and y in G. The product of any two elements in H is uniquely determined by this definition. Since $f(x, y) = 1$ for x in $Z(G)$, this correspondence ϕ between the group G and the multiplicative manifold H satisfies condition (1); and it satisfies (2), since $f(x^i, x^j) = 1$ for x in G and any integers i and j. This implies in particular that the picture of the group unit in G is the uniquely determined unit in H, and that the inverse of any element in H is uniquely determined. That the multiplication in H is commutative follows from

$$x^\phi y^\phi = (f(y, x)xy)^\phi = (f(y, x)(x, y)yx)^\phi = (f(x, y)yx)^\phi = y^\phi x^\phi;$$

and that the multiplication in H is associative, follows from

$$\begin{aligned}(x^\phi y^\phi)z^\phi &= (f(y, x)xy)^\phi z^\phi = (f(z, f(y, x)xy)f(y, x)xyz)^\phi \\ &= f(y, x)^\phi f(z, x)^\phi f(z, y)^\phi (xyz)^\phi\end{aligned}$$

and

$$\begin{aligned}x^\phi(y^\phi z^\phi) &= x^\phi(f(z, y)yz)^\phi = (f(f(z, y)yz, x)xf(y, z)yz)^\phi \\ &= f(y, z)^\phi f(z, x)^\phi f(y, x)^\phi (xyz)^\phi.\end{aligned}$$

Consequently H is an abelian group and ϕ satisfies (3), since all the elements $f(x, y)$ are contained in S.

† The idea for this proof has been suggested by an argument used by C. Hopkins for the proof of a similar theorem, cf. C. Hopkins, these Transactions, vol. 37 (1935), pp. 169–170.

Suppose finally that γ is an automorphism of G such that $S^\gamma \leqq S$. Then

$$(f(x, y)^\gamma)^2 = (f(x, y)^2)^\gamma = (x, y)^\gamma = (x^\gamma, y^\gamma) = f(x^\gamma, y^\gamma)^2,$$

consequently

$$f(x, y)^\gamma = f(x^\gamma, y^\gamma)$$

since $f(x, y)^\gamma$ is an element of S. Hence

$$\begin{aligned}(x^\phi y^\phi)^{\phi^{-1}\gamma\phi} &= (f(y, x)xy)^{\gamma\phi} = (f(y, x)^\gamma x^\gamma y^\gamma)^\phi \\ &= (f(y^\gamma, x^\gamma)x^\gamma y^\gamma)^\phi = x^{\gamma\phi}y^{\gamma\phi} \\ &= (x^\phi)^{\phi^{-1}\gamma\phi}(y^\phi)^{\phi^{-1}\gamma\phi};\end{aligned}$$

and this completes the proof of the Theorem.

Example B.2. This example is to show that the Theorem B.1 would not hold without the assumptions concerning the existence of a subgroup S without elements of order 2 such that $C(G) \leqq S^2 \leqq S$.

Denote by Z an infinite cyclic group generated by an element z, and by G^* a direct product of two infinite cyclic groups. If u^*, v^* is any basis of G^*, then an operation $x^* \circ y^*$ of G^* in Z which satisfies the conditions (1) to (4) of Corollary 2.3 is characterized by $u^* \circ v^* = z$. Let G be the (essentially uniquely determined) group such that $Z = C(G) = Z(G)$, $G^* = G/Z$ which realizes the operation $x^* \circ y^*$.

This group is conformal to a direct product of three infinite cyclic groups.

Suppose that ϕ is a one-one correspondence which maps G upon some group H and which satisfies the conditions:

(i) $(xy)^\phi = x^\phi y^\phi$ for x in $Z(G)$ and y in G.

(ii) $(x^n)^\phi = (x^\phi)^n$ for x in G and any integer n.

(iii) ϕ induces a homomorphism of G upon $H/Z(G)^\phi$.

(iv) If γ is a proper automorphism of G, then $\phi^{-1}\gamma\phi$ is an automorphism of H.

Note that (i) and (ii) are identical with the conditions (1) and (2) of the Theorem B.1, whereas (iii) and (iv) are even weaker conditions than the corresponding conditions (3) and (4) of the Theorem B.1.

Since ϕ is a one-one correspondence, an element $f(x, y)$ is uniquely determined by the equation

$$x^\phi y^\phi = (f(y, x)xy)^\phi$$

for x and y in G, and it is a consequence of (iii) that $f(y, x)$ is an element of $Z(G)$. It is a consequence of (i) that $f(x, y) = f(Z(G)x, Z(G)y)$ is independent of the choice of x and y in their respective classes in $G/Z(G)$; and it is a consequence of (ii) that $f(x^i, x^j) = 1$ for x in G and integral i and j.

Let now u and v be some representatives of the classes u^* and v^* of G/Z, respectively. Then an automorphism γ of G is defined by

$$u^\gamma = uv, \qquad v^\gamma = v^{-1}, \qquad z^\gamma = z^{-1},$$

since $(u^\gamma, v^\gamma) = (uv, v^{-1}) = (v, u) = z^{-1} = z^\gamma = (u, v)^\gamma$. It is therefore a consequence of condition (iv) that

$$\begin{aligned}(f(v, u)^\gamma u^\gamma v^\gamma)^\phi &= (f(v, u)uv)^{\gamma\phi} = (f(v, u)uv)^{\phi\phi^{-1}\gamma\phi} \\ &= (u^\phi v^\phi)^{\phi^{-1}\gamma\phi} = u^{\phi\phi^{-1}\gamma\phi} v^{\phi\phi^{-1}\gamma\phi} = u^{\gamma\phi} v^{\gamma\phi} \\ &= (u^\gamma)^\phi (v^\gamma)^\phi = (f(v^\gamma, u^\gamma) u^\gamma v^\gamma)^\phi;\end{aligned}$$

and this implies, since $f(u, v)$ is an element of Z, that

$$f(v, u)^{-1} = f(v, u)^\gamma = f(v^\gamma, u^\gamma) = f(v^{-1}, uv).$$

Another automorphism ι of G is defined by

$$u^\iota = v, \qquad v^\iota = u, \qquad z^\iota = z^{-1};$$

and on applying a similar argument it follows that $f(v, u)^{-1} = f(u, v)$.

Finally it is a consequence of all these equations and (ii) that

$$\begin{aligned}(v^\phi, u^\phi) &= v^\phi u^\phi (v^\phi)^{-1} (u^\phi)^{-1} = v^\phi u^\phi (v^{-1})^\phi (u^{-1})^\phi \\ &= f(u, v)^\phi f(v^{-1}, vu)^\phi f(u^{-1}, vuv^{-1})^\phi (v, u)^\phi \\ &= f(u, v)^\phi f(v^{-1}, uv)^\phi f(u^{-1}, u)^\phi (v, u)^\phi \\ &= f(u, v)^\phi f(v, u)^{-1\phi} (v, u)^\phi = (f(u, v)^2 (v, u))^\phi;\end{aligned}$$

and this last expression is not equal to one, since $(v, u) = z^{-1}$ is not the square of an element in Z. Thus H is not an abelian group.

If Z is a cyclic group of order 2^n for $0 < n$, z an element generating Z, if G^* is a direct product of two cyclic groups of order 2^{n+1}, and if u^*, v^* is a basis of G^*, then let G be the group which satisfies $C(G) = Z \leqq Z(G)$, $G/Z = G^*$, and which realizes the operations characterized by $u^* \circ v^* = z$, $P(2^{n+1}, u^*) = P(2^{n+1}, v^*) = 1$. (This group G has been discussed at the end of §8.) G is conformal to an abelian group, but practically the same argument as the one used above proves that there does not exist a transformation ϕ into an abelian group which satisfies the conditions (i) to (iv).

COROLLARY B.3. *If G is a group with abelian central quotient group, and if $C(G) = C(G)^2$ does not contain elements of order 2, then there exists a conformality ϕ of G upon an abelian group H which satisfies the conditions* (1) *and* (2) *of Theorem* B.1 *and transforms the (proper or improper) automorphisms of G into automorphisms of H.*

This is a consequence of the fact that every automorphism of G maps $C(G)$ upon a subgroup of $C(G)$ and of Theorem B.1.

Note that the conditions of this Corollary B.3 are satisfied, if $G/Z(G)$ is an abelian group whose elements are of finite odd order.

Appendix C. Operator groups. If G is a group, then Φ is said to be a *set of operators for* G, if x^ϕ is, for every x in G and for every ϕ in Φ, a uniquely determined element in G such that $(xy)^\phi = x^\phi y^\phi$. Φ may be called an *associative set of operators for the group* G, if a multiplication of the elements in Φ is defined which satisfies $(x^\phi)^\gamma = x^{\phi\gamma}$.

If G is a group, and if Φ is a set of operators for the group G, then the subgroup S of G is said to be Φ-*admissible* if $S^\phi \leqq S$ for every ϕ in Φ. The commutator group $C(G)$ is Φ-admissible for every set Φ of operators. But it is easy to construct groups G and sets Φ of operators of G such that the central $Z(G)$ is not Φ-admissible.

If G is a group, and if Φ is a set of operators for G, then the Φ-*central* $Z(G, \Phi)$ of G may be defined as the set of all those elements x in G which satisfy $xy = yx$, for every y in G, and $x^\phi = x$, for every ϕ in F; and the subgroup $C(G, \Phi)$ of G, which is generated by the elements (x, y) and $x^\phi x^{-1}$ for x, y in G and ϕ in Φ, may be called the Φ-*commutator group of* G. Both $Z(G, \Phi)$ and $C(G, \Phi)$ are normal Φ-admissible subgroups of G. $Z(G, \Phi)$ is the greatest subgroup of G in which the inner automorphisms of G and the operators in Φ induce the identity transformation. $C(G, \Phi)$ is the smallest subgroup of G whose quotient group is abelian and in whose quotient group all the operators in Φ induce the identity only. Thus $Z(G, \Phi)$ and $C(G, \Phi)$ are Φ-characteristic subgroups of the Φ-operator group G.

The obvious generalization of the groups with abelian central quotient group are the groups satisfying the equation

$$C(G, \Phi) \leqq Z(G, \Phi).$$

They are groups with abelian central quotient group, since $C(G) \leqq C(G, \Phi)$ and $Z(G, \Phi) \leqq Z(G)$. Thus it is advisable to consider G as an extension of a group S between $C(G, \Phi)$ and $Z(G, \Phi)$ (S is then also situated between $C(G)$ and $Z(G)$) by the group $G/S = G^*$. S and G^* are abelian groups essentially without operators. Hence the only problem is to characterize the operators in Φ by relations between G^* and S, and this may be done by practically the same method as has been used in §10 in order to describe the automorphisms in the group $\Omega(S < G)$ since the elements in Φ induce automorphisms of G which belong (as automorphisms) to $\Omega(S < G)$.

THE UNIVERSITY OF ILLINOIS,
URBANA, ILL.

GROUPS WITH PREASSIGNED CENTRAL AND CENTRAL QUOTIENT GROUP*

BY

REINHOLD BAER

Every group G determines two important structural invariants, namely its central $C(G)$ and its central quotient group $Q(G)=G/C(G)$. Concerning these two invariants the following two problems seem to be most elementary. If A and B are two groups, to find necessary and sufficient conditions for the existence of a group G such that A and $C(G)$, B and $Q(G)$ are isomorphic (existence problem) and to find necessary and sufficient conditions for the existence of an isomorphism between any two groups G' and G'' such that the groups A, $C(G')$ and $C(G'')$ as well as the groups B, $Q(G')$ and $Q(G'')$ are isomorphic (uniqueness problem). This paper presents a solution of the existence problem under the hypothesis that B (the presumptive central quotient group) is a direct product of (a finite or infinite number of finite or infinite) cyclic groups whereas a solution of the uniqueness problem is given only under the hypothesis that B is an abelian group with a finite number of generators.

There is hardly any previous work concerning these problems. Only those abelian groups with a finite number of generators which are central quotient groups of suitable groups have been characterized before.†

1. Before enunciating our principal results (in §2) some notation and concepts concerning abelian groups will have to be recorded for future reference.

If G is any abelian group, then the composition of the elements in G is denoted as addition: $x+y$. If n is any positive integer, then nG consists of all the elements nx for x in G, and G_n consists of all the elements x in G such that $nx=0$. $F(G)$ is the subgroup of all the elements of finite order in G, that is, the join of the groups G_n, and $F(G, p)$ is the subgroup of all those elements in $F(G)$ whose order is a power of the prime number p; in other words, $F(G, p)$ is the join of all the groups G_{p^n}. It is well known that $F(G)$ is the direct sum of the groups $F(G, p)$.

A set B of elements in G is termed *independent*, if all the elements in B are non-zero, if none of the elements in B is a multiple of another element in B, and if the group, generated by the elements in B, is the direct sum of the cyclic groups which are generated by the elements in B. If G is generated by

* Presented to the Society, September 10, 1937; received by the editors September 20, 1937.

† R. Baer, Mathematische Zeitschrift, vol. 38 (1934), p. 406.

B, and if B is independent, then B is called a *basis* of G. It may be mentioned that the elements u and v (neither zero) are dependent if there exist integers i and k such that $iu = kv \neq 0$.

If G is an abelian group, then there need not exist a basis of G, but there always exists a greatest independent subset of G. Each greatest independent subset of G contains a certain (finite or infinite) number of elements, and the least of these numbers is called the *rank* $r(G)$ of G. If either every non-zero element in G is of infinite order or every non-zero element in G is of order p, then every greatest independent subset of G contains exactly $r(G)$ elements.

Numerical invariants:

$$r(G, 0) = r(G \pmod{F(G)}), \qquad r(G, p^i) = r((p^{i-1}G)_p \pmod{(p^iG)_p}).$$

If G is a direct sum of cyclic groups, then in every decomposition of G into indecomposable direct summands there are exactly $r(G, 0)$ cyclic direct summands of infinite order and exactly $r(G, p^i)$ cyclic direct summands of order p^i. It is a consequence of this fact that the structure of direct sums of cyclic groups is completely determined by the invariants $r(G, 0)$, $r(G, p^i)$.

If S is any subgroup of the abelian group G, then G/S denotes the group of classes of residues of G (mod S), and the direct sum of the abelian groups $G, H, \cdots, K, \cdots$ is denoted by $G+H+\cdots+K+\cdots$.

2. The following theorems contain the main results of this investigation:

EXISTENCE THEOREM. *If V is an abelian group and if G is a direct sum of cyclic groups, then the following conditions are necessary and sufficient for the existence of a group whose central is V and whose central quotient group is isomorphic to G:*

(i) *If G contains elements of order p^i, then V contains elements of order p^i.*

(ii) *If G contains elements of infinite order, then V contains elements of infinite order, or the orders of the elements in $F(V)$ are not bounded.*

(iii) *If the orders of the elements in $F(G)$ are bounded, and if $r(G, 0)$ is a finite positive integer, then V contains elements of infinite order and $1 < r(G, 0)$.*

(iv) *If the orders of the elements in $F(G)$ are bounded, and if $r(G, 0)$ is an odd (finite) number, then V contains two independent elements of infinite order $(1 < r(V, 0))$.*

(v) *If $G = F(G)$, $F(G, p) \neq 0$, and the orders of the elements in $F(G, p)$ are bounded, $F(G, p)$ contains at least two independent elements of maximum order.*

(vi) *If $G = F(G)$, if the orders of the elements in $F(G, p)$ are bounded, if $r(G, p^{i+k})$ is finite for $0 \leq k$, and if $r(G, p^i)$ is odd, then V contains two independent elements of order p^i, $(1 < r((p^{i-1}V)_p)$.*

The following corollary is easily derived from this theorem:

COROLLARY. *The direct sum G of cyclic groups is isomorphic to the central quotient group of a suitable group if, and only if,*

(a) $r(G, 0)=1$ *implies that the orders of the elements in* $F(G)$ *are not bounded;*

(b) $G=F(G)$ *and* $r(G, p^i)=1$ *imply that* G *contains elements of order* p^{i+1}.

For (a) and (b) are part of the conditions (iii) and (v), respectively, and if G satisfies the conditions (a) and (b), then it is fairly obvious how to construct a group V such that G and V satisfy the conditions (i) to (vi) of the existence theorem.

UNIQUENESS THEOREM. *If G and V are abelian groups, and if G may be generated by a finite number of its elements, then the following conditions are necessary and sufficient for the existence of one, and essentially only one, group whose central is V and whose central quotient group is isomorphic to G:*

(1) *If p is a prime number such that G contains elements of order p, then $V=pV$ and V contains elements of order p.*

(2) *If G contains elements of infinite order, then $V=nV$ for every positive integer n, and V contains elements of infinite order.*

(3) *If $r(G, 0)\neq 0$, then $1<r(G, 0)<4$.*

(4) *If $r(G, 0)=3$, then $F(G)=0$, $F(V)=0$, and $2=r(V, 0)$ $(=r(V))$.*

(5) *If $r(G, 0)=2$ and $F(G, p)\neq 0$, then $F(G, p)$ is a direct sum of an odd number of isomorphic cyclic groups, and $1=r(V_p)$.*

(6) *If $G=F(G)$ (that is, if G is a finite group), if $F(G, p)\neq 0$, and if $2<r(V_p)$, then $F(G, p)$ is a direct sum of two isomorphic cyclic groups.*

(7) *If $G=F(G)$, $F(G, p)\neq 0$, and $2=r(V_p)$, then $F(G, p)$ is a direct sum of two cyclic groups of order p^{i+j} and one cyclic group of order p^i where $0\leqq i$, $0\leqq j$, $0<i+j$.*

(8) *If $G=F(G)$, $F(G, p)\neq 0$, and $1=r(V_p)$, then $F(G, p)$ is a direct sum of two isomorphic groups.*

Remark. If V and G satisfy condition (1), then at least one of them is infinite.

The proofs of these theorems will occupy us throughout this paper. In fact we shall prove slightly more than is needed for these theorems. Thus it will be possible to derive, from the facts presented in the following sections, the following statement which shows that the hypothesis of the existence of a finite basis in G is not a necessary one (though it is needed):

If G is a direct sum of finite cyclic groups of order n, if $r(G)=\aleph_0$, and if $V=nV$ and V_n is a cyclic group of order n, then there exists one, and essentially only one, group whose central is V and whose central quotient group is isomorphic to G.

3. The proofs of the theorems enunciated in §2 are based on the following theorems which transform the "metabelian" problems into abelian problems and which have been proved by the author in an earlier paper.* The following definitions are found necessary for the statement of these transformation theorems:

If G and V are two abelian groups, then the operation xy is called a *multiplication of G in V* if it obeys the following rules:

(3.1) *If x and y are elements in G, then xy is a uniquely determined element in V.*

(3.2) $0 = xx = xy + yx,\ x(y+z) = xy + xz.$

A multiplication xy of G in V may be called a *proper multiplication of G in V*, if it satisfies the condition:

(3.3) *$wx = 0$ for every x in G if, and only if, $w = 0$.*

We define *an admissible set of functions of G in V* to consist of a proper multiplication xy of G in V and of functions $P(n, x)$ which satisfy the following conditions:

(3.4) *If G contains elements of order n, and if x is an element in G_n, then $P(n, x)$ is a uniquely determined element in V* (mod nV).

(3.5) *If G contains elements of order n, and if x and y are elements in G_n, then*

$$P(n, x + y) = n(n-1)2^{-1}xy + P(n, x) + P(n, y).$$

(3.6) *If m is a positive integer, G contains elements of order nm, and if x is an element in G_n, then*

$$P(nm, x) = mP(n, x).$$

(3.7) *If G contains elements of order nm and if x is an element in G_{nm}, then*

$$P(n, mx) = nV + P(nm, x).$$

The following theorem constitutes a transformation of the existence theorem:

THEOREM.† *There exists a proper multiplication of the abelian group G in the abelian group V if, and only if, there exists a group whose central is isomorphic to V and whose central quotient group is isomorphic to G.*

The following is a transformation of the uniqueness theorem:

* R. Baer, *Groups with abelian central quotient group*, these Transactions, vol. 44 (1938), pp. 357–386.

† Baer, ibid., Corollary 2.3.

THEOREM.* *Any two groups whose centrals are isomorphic to the abelian group V and whose central quotient groups are isomorphic to the direct sum G of cyclic groups are isomorphic if, and only if, there exists for any pair*

$$xy,\ P(n, x) \text{ and } x \circ y,\ P_0(n, x)$$

of sets of admissible functions of G in V an automorphism ϕ of V and an automorphism γ of G such that

$$(xy)^\phi = x^\gamma \circ y^\gamma \text{ and } P(n, x)^\phi = P_0(n, x^\gamma).$$

4. The relations of the conditions which are involved in the definition of a multiplication xy of the abelian group G in the abelian group V may be analyzed as follows:

If $0=xy+yx$ is always satisfied, then $2xx=0$ is inferred by putting $x=y$. If $xx=0$ and $x(y+z)=xy+xz$, as well as $(y+z)x=yx+zx$ (this would be a consequence of $xy+yx=0$!), are satisfied, then $0=(x+y)(x+y)=xx+xy+yx+yy=xy+yx$.

If, finally, xy is a proper multiplication of G in V, then the hypothesis that G is an abelian group is a consequence of the other conditions.

(4.1) *If xy is a multiplication of G in V, and if u is an element in G_n, then ux and xu are elements in V_n for every x in G.*

For multiplications are associative with regard to multiplication by integers, that is, $n(xy)=(nx)y=x(ny)$.

If xy is a multiplication of G in V, then to every element g in G there corresponds the homomorphism of G into V which is defined by mapping the element x in G upon the element gx in V. To the sum of two elements in G there corresponds the sum of the corresponding homomorphisms, and to any two different elements in G there correspond different homomorphisms of G into V if, and only if, xy is a proper multiplication of G in V. Thus every multiplication of G in V defines a representation of G as a group of homomorphisms of G into V, and this representation is a true one if, and only if, the multiplication is a proper one.

5. We prove the following statement:

(5.1) *If xy is a multiplication of G in V, and if u and v are elements in G such that u and uv have the same order, then u and v are independent elements in G.*

Proof. Suppose that h and k are integers such that $hu=kv$. If neither h nor k is 0, then let d be the g.c.d. of h and k. There exist, therefore, integers

* Baer, ibid., Theorem 6.2.

h', k' such that $d=hh'+kk'$, and it follows that

$$d(uv)=hh'uv+kk'uv=h'(hu)v+k'u(kv)=h'kvv+k'huu=0.$$

Hence d is a multiple of the order of uv, and, since u and uv have the same order, this implies $du=0$ and therefore $hu=0$ provided u and uv are of finite order; whereas the above equation leads to a contradiction if u and uv are both of infinite order. Thus u and v are independent.

(5.2) *If xy is a proper multiplication of G in V, then there exists corresponding to every element u in $F(G)$ an element u' in G such that u and uu' have the same order; and if u is an element of infinite order such that ux is an element of finite order for every x in G, then the orders of the elements ux for x in G are not bounded.*

Proof. Suppose that all the elements ux for x in G are of finite order and that the orders of the elements ux for x in G are bounded. Then the l.c.m. of the orders of the elements ux is a finite positive integer m. It follows that $0=m(ux)=(mu)x$ for every x in G, consequently $mu=0$ by (3.3). Clearly m is the order of u, as follows from (4.1) and the definition of m. Since it is fairly clear how to prove the existence of an element u' in G such that m is the order of uu', this completes the proof.

(5.3) *If xy is a multiplication of G in V, and if b' and $b'b''$ have the same order p^n, then $b''\neq 0 \pmod{pG}$.*

Proof. If $b''=pb$, then $b'b''=p(b'b)$; consequently $p^{n-1}(b'b'')=(p^nb')b=0$.

6. If xy is a multiplication of G in V, and if S is a subgroup of G and T is a subgroup of V, then xy induces a multiplication of S in V/T. But clearly this induced multiplication need not be a proper multiplication of S in V/T, even if the inducing multiplication is a proper one.

(6.1) *If xy is a proper multiplication of G in V, if m is a positive integer such that mG contains multiples, not zero, of every non-zero element in G, and if T is a subgroup of V such that 0 is the cross cut of T and mV, then xy induces a proper multiplication of G in V/T.*

Proof. If $w\neq 0$ is an element in G, then there exists an integer k and an element u in G such that $kw=mu\neq 0$. Since xy is a proper multiplication of G in V, there exists an element w' in G such that $0\neq kww'=muw'$. Since 0 is the only element contained in T as well as in mV, it follows that kww', and consequently ww', are not elements in T; and this implies that xy defines a proper multiplication of G in V/T.

(6.2) *If xy is a proper multiplication of G in V, if G is generated by its subgroups G_1 and G_2, and if there exists a positive integer m such that $mx_1x_2=0$ for x_i in G_i and such that mG_1 contains multiples, not zero, of every non-zero element in G_1, then xy defines a proper multiplication of G_1 in V.*

Proof. If $w\neq 0$ is an element in G_1, then there exists an integer k and an element w' in G_1 such that $kw=mw'\neq 0$. Since xy is a proper multiplication of G in V, there exists an element u in G such that $kwu\neq 0$. Since G is generated by G_1 and G_2, there exist elements u_i in G_i such that $u=u_1+u_2$. Consequently, $0\neq kwu=mw'(u_1+u_2)=kwu_1+mw'u_2=kwu_1$, since w' is an element in G_1. Hence $wu_1\neq 0$; and xy defines, therefore, a proper multiplication of G_1 in V.

Note that the hypothesis $mx_1x_2=0$, for x_i in G_1, is satisfied if $mG_2=0$.

7. If G is the direct sum of its subgroups G_v, and if, for every v, a multiplication xy of G_v in V is given, then there exists one and only one multiplication xy of G in V which induces the given multiplications in the groups G_v and satisfies $rs=0$ for elements r and s which belong to different components G_v. This multiplication xy of G in V is a proper one if, and only if, the induced multiplications of the groups G_v in V are proper multiplications.

Notation. If xy is a multiplication of G in V, then $M(G, xy)$ is the subgroup of V generated by the elements xy for x and y in G; and if B is a subgroup of G, and S is a subgroup of V, then $(B<G; S, xy)$ consists of all the elements w in G such that $wb\equiv 0 \pmod{S}$ for every b in B.

Thus a multiplication xy of G in V is also a multiplication of G in $M(G, xy)$, and a proper multiplication of G in V is a proper multiplication of G in $M(G, xy)$. (If G is generated by two elements, then $M(G, xy)$ is a cyclic group (which may be 0); and if G is generated by a finite number of elements, then so is $M(G, xy)$.) It may be noted that $(B<G; S, xy)$ is always a subgroup of the group G.

LEMMA 7.1. *If xy is a multiplication of G in V, if B is a subgroup of G generated by the elements b' and b'' such that b', b'', and $b'b''$ all have the same order, and if $M(G, xy)$ is the direct sum of $M(B, xy)$ and its subgroup S, then G is the direct sum of $(B<G; S, xy)$ and the cyclic groups, generated by b' and b''.*

Proof. It is a consequence of (5.1) that b' and b'' form a basis of B, since b', b'', and $b'b''$ all have the same order. If w is an element in the cross cut of B and $(B<G; S, xy)$, then $wb=0$ for every element b in B, since 0 is the cross cut of S and $M(B, xy)$. Since xy defines a proper multiplication of B in V, this implies that $w=0$. If finally u is any element in G, then $ub'=s'+r'(b'b'')$ and $ub''=s''+r''(b'b'')$, for s', s'' in S and r', r'' suitable integers. Hence $u-r''b'+r'b''$ is an element in $(B<G; S, xy)$, and G is therefore the direct

sum of B and $(B<G; S, xy)$. This completes the proof.

COROLLARY 7.2. *If xy is a proper multiplication of the group G with a finite number of generators in the cyclic group $M(G, xy)$, then*

(a) *G is either a direct sum of a finite number of infinite cyclic groups or a direct sum of a finite number of finite cyclic groups; and*

(b) *there exists a basis $b_1', b_1'', \cdots, b_k', b_k''$ of G with the following properties:*

(i) *b_j', b_j'', and $b_j' b_j''$ all have the same order;*

(ii) *$b_i' b_j' = b_i'' b_j'' = b_i' b_k'' = 0$ for $i \neq h$;*

(iii) *$b_j' b_j'' = r_j b_{j-1}' b_{j-1}''$ for $1 < j \leqq k$; and if G is finite, then r_j is a divisor of the order of b_{j-1}'.*

(c) *G is a direct sum of two isomorphic groups G' and G'' such that $M(G', xy) = M(G'', xy) = 0$.*

Proof. (c) is an obvious consequence of (b), and (a) follows easily from (5.2) and the fact that the non-zero elements in a cyclic group are either all of finite order or all of infinite order. Since $M(G, xy)$ is a cyclic group, there exists a pair of elements b_1', b_1'' in G such that $b_1' b_1''$ generates $M(G, xy)$. Since $b_1' b_1''$ is an element of maximum order in $M(G, xy)$, it follows from (5.2) that b_1', b_1'', and $b_1' b_1''$ all have the same order. If B is the subgroup of G, generated by b_1' and b_1'', then it follows from $M(G, xy) = M(B, xy)$ and Lemma 7.1 that G is the direct sum of $(B<G; 0, xy)$ and the cyclic groups generated by b_1' and b_1''. Since $uv = 0$ for u in $(B<G; 0, xy)$ and v in B, it follows that xy defines proper multiplications of B as well as of $(B<G; 0, xy)$. Since $M((B<G; 0, xy), xy)$ is a subgroup of the cyclic group $M(G, xy)$, it is itself cyclic and (b) of Corollary 7.2 may be applied to $(B<G; 0, xy)$ since it is generated by less elements than G. Now the proof is easily completed by complete induction.

COROLLARY 7.3. *If G is a direct sum of $\aleph_0$ cyclic groups of equal finite order, if xy is a proper multiplication of G in V, and if $M(G, xy)$ is a cyclic group, then there exists a basis b_j', b_j'' for $j = 1, 2, \cdots$ of G with the following properties:*

(i) *b_j', b_j'', and $b_j' b_j''$ all have the same order.*

(ii) *$b_i' b_j' = b_i'' b_j'' = b_k' b_k'' = 0$ for $h \neq k$.*

(iii) *$b_1' b_1'' = \cdots = b_i' b_i'' = \cdots$ generates $M(G, xy)$.*

Proof. Let $g_1, g_2, \cdots, g_i, \cdots$ be an enumeration of the elements in G. Then by complete induction elements b_j', b_j'' and groups G_j will be defined for $0 < j$ with the following properties:

(1) The group B_i which is generated by the elements b_j', b_j'' with $j < i$ has these elements as a basis.

(2) The elements b_j', b_j'' satisfy (i) to (iii).
(3) B_i contains the elements g_j with $j<i$.
(4) G is the direct sum of G_i and B_i.
(5) $uv=0$, if u in G_i and v in B_i.
(6) xy induces proper multiplications of B_i and of G_i in V.

Since $B_1=0$, $G_1=G$ is a suitable start of this construction, it may be assumed that elements b_j', b_j'', for $j<i$ and a group G_i, have been defined which meet the requirements (1) to (6). Then $g_i=c+d$ with c in B_i and d in G_i. Since G_i is by condition (4) a direct sum of $\aleph_0$ cyclic groups of equal order, d is a multiple of an element b_i' in G_i which is of maximum order in G. There exists, by (5.2) and by the fact that $M(G_i, xy)$ is a cyclic group, an element b_i'' in G_i such that b_i', b_i'', and $b_i'b_i''$ have the same order, and such that $b_i'b_i''$ generates $M(G_i, xy)$. Since the orders of the elements in G are finite, b_i'' may be chosen in such a way that $b_i'b_i''=b_1'b_1''$. Finally we may put $G_{i+1}=(C<G_i; 0, xy)$, where C is the subgroup generated by b_i', b_i''. Then it follows from Lemma 7.1 that the elements b_j', b_j'', for $j<i+1$, and the group G_{i+1} satisfy (1) to (6).

The subgroup generated by all the elements b_i', b_i'', for $i=1, 2, \cdots$, contains every element in G by (3); and the b_i', b_i'' form therefore, by (1), a basis of G which satisfies (i) to (iii) by (2).

COROLLARY 7.4. *If G is a direct sum of two infinite cyclic groups and of an odd number of cyclic groups of the same finite order n, if xy is a proper multiplication of G in V such that $F(M(G, xy))$ is a cyclic group, then there exists a basis $b, b', b'', b_1', b_1'', \cdots, b_k', b_k''$ of G with the following properties:*

(i) $bb'=b_1'b_1''=\cdots=b_k'b_k''=c$.
(ii) *c, b, b_j', and b_j'' are elements of order n.*
(iii) *b', b'', and $b'b''$ are elements of infinite order.*
(iv) $bb''=bb_j'=bb_j''=b'b_j'=b'b_j''=b''b_j'=b''b_j''=b_i'b_j'=b_i''b_j''=b_j'b_k''=0$ *for $j\neq h$.*

Proof. Denote by N the set $(F(G)<F(G); 0, xy)$ of all elements w in $F(G)$ such that $wx=0$ for every x in $F(G)$. Then xy defines a proper multiplication of $F(G)/N$ in the cyclic group $F(M(G, xy))$. Therefore there exists, by Corollary 7.2, a basis $b_1', b_1'', \cdots, b_k', b_k''$ of $F(G)$ (mod N) which satisfies the conditions (i) to (iii) of (b) of Corollary 7.2.

Denote by u_1, u_2 any pair of elements in G which forms a basis of G (mod $F(G)$), and let w_j, for $j=1, 2, 3$ be any three elements in N whose order is a prime number p (dividing n). If d is some element of order p in $M(G, xy)$, then $w_ju_i=r_{ji}d$. If h_1, h_2, and h_3 are not trivial solutions of the two congruences in three unknowns $\sum_{j=1}^{3}h_jr_{ji}\equiv 0 \pmod p$, then $w=\sum_{j=1}^{3}h_jw_j$ belongs to N and

satisfies $wu_i=0$. Thus $w=0$, since xy is a proper multiplication of G in V. Hence the rank of N_p is at most two, and this implies, together with the existence of the particular basis b_j', b_j'' of $F(G)$ (mod N) (mentioned before), that N_p is a cyclic group. N is therefore a cyclic group of order n, and it may be assumed, without loss in generality, that the b_j', b_j'' satisfy (i). If g generates N, then $gu_i=r_ic$, where c generates $F(M(G, xy))$. If r is the g.c.d. of r_1 and r_2, then there exist integers r_1' and r_2' such that $r_1r_1'+r_2r_2'=r$, and it follows from (5.2) that r and n are relatively prime. (If $r_2=0$, then r_1 is relatively prime to n, and g may be chosen in such a way that $r_1=1$.) The elements

$$u_1' = r_1' u_1 + r_2' u_2, \qquad u_2' = -r_2 r^{-1} u_1 + r_1 r^{-1} u_2$$

form a basis of G (mod $F(G)$) which satisfies $gu_1'=rc$, $gu_2'=0$; thus it may be assumed without loss in generality that an element b, generating N, and a basis v_1, v_2 of G (mod $F(G)$) have been chosen in such a way that $bv_1=c$, $bv_2=0$. Put finally

$$b_i = v_i + \sum_{j=1}^{k} (s_{ij}' b_j' + s_{ij}'' b_j''),$$

where the numbers s are determined as solutions of the equations

$$v_ib_j' = s_{ij}'' c, \qquad v_ib_j'' = -s_{ij}' c.$$

Then a basis of G has been found which meets the requirements of the Corollary 7.4.

8. We prove the following lemma:

LEMMA 8.1. *If G is a direct sum of three cyclic groups, if xy is a proper multiplication of G in V, and if $M(G, xy)$ is a direct sum of two cyclic groups, then there exists a basis b_1, b_2, b_3 of G such that $b_1b_3=0$, whereas b_1b_2 and b_2b_3 form a a basis of $M(G, xy)$.*

Proof. Denote by u, v a basis of $M(G, xy)$, by $\bar{u}$ the cyclic subgroup, generated by u, and by $\bar{v}$ the cyclic subgroup, generated by v. Then $M(G, xy)=\bar{u}+\bar{v}$. If both u and v are of finite order, it may be assumed that the order of u is a divisor of the order of v. (Then it may happen that $u=0$.) There exists a pair of elements b', b'' in G such that $v=b'b''$; and it is a consequence of (5.2) and the choice of v that b', b'', and v have the same order. If B is the subgroup of G generated by b' and b'', then it follows from Lemma 7.1 that b', b'' form a basis of B and that $G=B+(B<G; \bar{u}, xy)$. Clearly $(B<G; \bar{u}, xy)$ is a cyclic group generated by an element c. There exist in B elements b such that $bc=u$ and amongst these elements b there is one b_2 such that the homomorphism of B, induced by b_2, maps B upon $\bar{v}$. It is now clear how to complete the proof.

9. Two multiplications xy and $x \circ y$ of G in V are called *isomorphic*, if there exists an automorphism γ of G and an isomorphism μ of $M(G, xy)$ upon $M(G, x \circ y)$ such that

$$(xy)^{\mu} = x^{\gamma} \circ y^{\gamma} \text{ for } x \text{ and } y \text{ in } G.$$

If the isomorphism μ may be chosen in such a fashion that it is induced by some automorphism of V, then the two multiplications are termed *equivalent*.

The following propositions are fairly obvious consequences of the statements in §§7 and 8.

If G is a direct sum of three (but not of two) cyclic groups, then any two proper multiplications of G such that $M(G, \cdots)$ are direct sums of two cyclic groups are isomorphic.

If G is a direct sum of $\aleph_0$ isomorphic finite cyclic groups, or if G is a finite group, then any two proper multiplications with the cyclic $M(G, \cdots)$ are isomorphic.

If G is a direct sum of two infinite cyclic groups and of an odd finite number of isomorphic finite cyclic groups, then any two proper multiplications of G with cyclic $F(M(G, \cdots))$ are isomorphic.

It is a consequence of (5.2) that any two proper multiplications of a direct sum of two cyclic groups are isomorphic.

If G is a direct sum of a finite number of infinite cyclic groups, and if xy is a proper multiplication of G such that $M(G, xy)$ is a cyclic group, then the numbers r_i appearing in (iii) of (b) of Corollary 7.2 may be chosen as positive integers. Then it is possible to prove that they are invariants with regard to isomorphisms. (This proof follows from the consideration of the subgroup $(G<G; nM(G,xy), xy)$ of all elements w in G such that $wx \equiv 0 \pmod{nM(G, xy)}$ for every x in G.) It is then a consequence of Corollary 7.2 that two proper multiplications of G with cyclic $M(G, \cdots)$ are isomorphic if, and only if, they induce the same invariants r_i.

10. The object of the next three sections is to construct proper multiplications of somewhat elementary groups G. They will be combined afterwards to form proper multiplications of the more general types of groups G.

If B is a basis of the group G, then a multiplication xy of G in V is completely determined by the values of the products bb' for b and b' in B, and these values may be chosen completely at random with the two restrictions that $bb = 0$ and that $bb' = -b'b$. In enumerating the values of the products bb' it may be understood that the product bb' need not be mentioned, if the value of $b'b$ has been given, and that $bb' = 0$, if the value of neither bb' nor $b'b$ has been mentioned.

11. We prove the following proposition:

(11.1) *If G is a direct sum of finite cyclic groups, if G is a direct sum of two isomorphic groups, and if V contains elements of order n whenever G contains elements of order n, then there exists a proper multiplication of G in V.*

Proof. There exists a basis of G, consisting of pairs b_v', b_v'' such that b_v and b_v'' always have the same order. There exists in V an element c_v such that b_v', b_v'', and c_v have the same order. Then a proper multiplication of G in V is determined by the rule $b_v' b_v'' = c_v$, for every v.

(11.2) *If G is a direct sum of cyclic groups whose orders are powers of a fixed prime number p, the orders of the elements in G are not bounded, and V contains elements of every order p^i, then there exists a proper multiplication of G in V.*

Proof. $G=G'+G''+G'''$, where G' and G'' are isomorphic groups and where G''' has a basis $b_1, b_2, \cdots, b_i, \cdots$ such that the orders of the elements b_i are not bounded and such that the order of b_{i-1} divides the order of b_i. There exists in V an element c_i whose order equals the order of b_i. A proper multiplication of G''' in V is characterized by the equations $b_i b_{i+1}=c_i$, for $i=1, 2, \cdots$. By (11.1) there exists a proper multiplication of $G'+G''$ in V, consequently there exists a proper multiplication of G in V.

(11.3) *If G is a direct sum of two cyclic groups of order p^n and of a finite number of cyclic groups whose orders are divisors of p^n, and if V contains elements of order p^n and contains two independent elements of order p^i in case $r(G, p^i)$ is odd, then there exists a proper multiplication of G in V.*

Proof. This is a consequence of (11.1) if all the numbers $r(G, p^i)$ are even. If not all the numbers $r(G, p^i)$ are even, then there exists a greatest integer m such that $r(G, p^m)$ is odd and $0<m\leqq n$. V contains an element u of order p^n and an element v of order p^m which is independent of u. There exists a basis $b_1, b_2, \cdots$ of G such that the order $p^{n(j)}$ of b_j satisfies the inequality $n(j)\leqq n(j-1)$ and consequently $n=n(1)=n(2)$. If $n=m$, then b_3 is of order p^n, and in this case we put $h=3$. If $m<n$, then let h be the smallest number such that $n(h)=m$. The number h thus defined is in both cases an odd integer and $n(2j-1)=n(2j)$, for $0<2j<h$.

It is now easily verified that a proper multiplication xy of G in V is characterized by the equations

$$b_{2j-1}b_{2j} = p^{n-n(2j)}u, \quad \text{for} \quad 0 < 2j < h,$$
$$b_{j-1}b_j = p^{m-n(j)}v, \quad \text{for} \quad h \leqq j.$$

(11.4) *If G is a direct sum of an infinity of cyclic groups of order p^n and of cyclic groups whose orders divide p^n, and if V contains elements of order p^n, then there exists a proper multiplication of G in V.*

Proof. There exists a basis $b_1, b_2, \cdots, b_v, \cdots$ of G which is well ordered by the subscripts of its elements b_v in such a fashion that there is no last element in the basis and that the order $p^{n(v)}$ of b_v divides $p^{n(k)}$ for $v<k$. There exists an element u of order p^n in V.

A proper multiplication of G in V is characterized by the equations

$$b_v b_{v+1} = p^{n-n(v)}u.$$

(11.5) *Suppose that G is a direct sum of two cyclic groups of order p^n and of a finite number of cyclic groups whose orders divide p^n and that V contains an element of order p^n and two independent elements of order p^i, if $r(G, p^i)$ is odd. Then there exist two non-isomorphic, proper multiplications of G in V if one of the following conditions is satisfied*:

(a) $2<r(G_p), 1<r(V_p)$, *and G is a direct sum of two isomorphic groups.*

(b) $2<r(G_p), 2<r(V_p)$.

(c) $3<r(G_p)$ *and G is not a direct sum of two isomorphic groups.*

Proof. If (a) is satisfied by G and by V, then there exists a basis $b_1, b_2, \cdots, b_{2k}$ such that $1<k$ and such that b_{2i-1} and b_{2i} have the same order $p^{n(i)}$. V contains an element u of order p^n and an element w of order p which is independent of u. If $h=0$ and if $h=1$, then a proper multiplication

$$x \underset{h}{\circ} y$$

of G in V is defined by the equations

$$b_{2i-1} \underset{h}{\circ} b_{2i} = p^{n-n(i)}u, \qquad b_{2i} \underset{h}{\circ} b_{2i+1} = hw$$

and

$$x \underset{0}{\circ} y \quad \text{and} \quad x \underset{1}{\circ} y$$

are clearly non-isomorphic multiplications of G in V.

If (b) is satisfied, then it may be assumed that G is not a direct sum of two isomorphic groups, since otherwise (a) might be applied. Then there exists a greatest integer m such that $r(G, p^m)$ is odd and $0<m\leqq n$. V contains a subgroup V' which is a direct sum of a cyclic group of order p^n and of a cyclic group of order p^m; and V contains an element w of order p which is not contained in V'.

There exists by (11.3) a proper multiplication xy of G in V', and it is a consequence of Corollary 7.2 that $M(G, xy)$ is not a cyclic subgroup of V'. Denote by $x \ominus y$ the multiplication of G in V which is characterized by the equations

$$b_1 \ominus b_k = w,$$

where b_1, b_k is a pair of elements in a basis of G such that $b_1 b_k = 0$. (That it is possible to determine xy and a basis of G in such a fashion may be verified by looking over the proof of (11.3).) Then

$$x \circ y = (xy) + (x \ominus y)$$

is a proper multiplication of G in V which satisfies

$$2 = r(M(G, xy)_p) < 3 = r(M(G, x \circ y)_p);$$

consequently xy and $x \circ y$ are non-isomorphic proper multiplications of G in V.

Assume now that the conditions of (c) are satisfied. Then there exists a greatest integer m such that $r(G, p^m)$ is odd and $0 < m \leqq n$. Furthermore V contains an element u of order p^n and an element v of order p^m which is independent of u.

It is finally possible to decompose G in the form

$$G = G' + G'' + \sum_{j=1}^{k} Z(j),$$

where G' and G'' are isomorphic groups and the maximum order of the elements in G' is p^n, and where $Z(j)$ is a cyclic group of order $p^{n(j)}$, $1 \leqq k$; $0 < n(k) < \cdots < n(1) = m \leqq n$.

Two cases may be distinguished.

Case 1. G' (and therefore G'') is a cyclic group of order p^n. Since G is not a direct sum of three cyclic groups, this implies that 1 is less than k. Denote by g', g'', and $z(j)$ elements which generate G', G'', and $Z(j)$, respectively.

A proper multiplication xy of G in V is characterized by the equations

$$g'g'' = u, \qquad g''z(1) = v, \qquad z(j-1)z(j) = p^{n-n(j)}u.$$

If $x = x'g' + x''g'' + \sum_{j=1}^{k} x(j)z(j)$ is any element in G, then

$$xg' = -x''u, \qquad xg'' = x'u - x(1)v, \qquad xz(1) = x''v - x(2)p^{n-n(2)}u,$$
$$xz(j) = (x(j-1)p^{n-n(j)} - x(j+1)p^{n-n(j+1)})u, \quad \text{for} \quad 1 < j < k,$$
$$xz(k) = x(k-1)p^{n-n(k)}u.$$

Denote by $W(xy, H)$ the set of all elements w in the subgroup H of G for which *the set wG of all the elements wx for x in G is a cyclic subgroup of V.* (This notation shall be used throughout this proof.) Clearly the element w in H belongs to $W(xy, H)$ if, and only if, the elements wg', wg'', $wz(j)$, for $1 \leqq j \leqq k$, generate a cyclic subgroup of V. It may now be computed that $W(xy, G)$ consists exactly of the elements of the form

$$x'g' + p^m x''g'' + \sum_{j=2}^{k} x(j)z(j)$$

and of the elements which may be represented in the form

$$x'g' + h\sum_{j=0}^{i} p^{h(j)}z(2j+1),$$

where $k=2i+1$, $h(j)=(n(2j+2)-n(2j+3))+\cdots+(n(k-1)-n(k))$, or the elements which may be represented in the form

$$x'g' + h\sum_{j=0}^{i-1} p^{h(j)}z(2j+1),$$

where $k=2i$, $h(j)=(n(2j+2)-n(2j+3))+\cdots+(n(k-2)-n(k-1))+n(k)$.

Another proper multiplication $x \circ y$ of G in V is characterized by the equations

$$g' \circ g'' = u, \qquad g'' \circ z(1) = v, \qquad g' \circ z(k) = p^{m-n(k)}v,$$
$$z(j-1) \circ z(j) = p^{n-n(j)}u, \quad \text{for} \quad 1 < j \leqq k.$$

Suppose now that $w=w'g'+w''g''+\sum_{j=1}^{k} w(k)z(k)$ is an element of G_p. Then

$$w \circ g' = -w''u - w(k)p^{m-n(k)}v, \qquad w \circ g'' = w'u - w(1)v,$$
$$w \circ z(1) = w''v - w(2)p^{n-n(2)}u,$$
$$w \circ z(j) = -w(j+1)p^{n-n(j+1)}u, \quad \text{for} \quad 1 < j < k,$$
$$w \circ z(k) = 0,$$

since $n(k)<n(k-1)\leqq m\leqq n$. Consequently $W(x \circ y, G_p)$ consists exactly of those elements which may be represented in the form

$$p^{n-1}w'g' + p^{n-1}w''g'' + \sum_{j=2}^{k-1} p^{n(j)-1}w(j)z(j) \quad \text{or} \quad p^{n-1}w'g' + p^{m-1}w(1)z(1),$$

if $m<n$, and which may be represented in the form

$$p^{n-1}w'g' + \sum_{j=2}^{k-1} p^{n(j)-1}w(j)z(j) \quad \text{or} \quad p^{n-1}w'g' + p^{m-1}w(1)z(1),$$

if $n=m$.

Thus $W(xy, G_p)$ and $W(x \circ y, G_p)$ are of essentially different structure, and the two proper multiplications xy and $x \circ y$ of G in V are not isomorphic.

Case 2. G' (and therefore G'') is not a cyclic group. Then there exists a basis $b_1', \cdots, b_i'$ of G', $b_1'', \cdots, b_i''$ of G'' such that $1<i$, b_j' and b_j'' have

the same order $p^{m(j)}$ and $m(i) \leqq \cdots \leqq m(1) = n$. Proper multiplications xy and $x \circ y$ of G in V are characterized by the equations

$$b_j' b_j' = p^{n-m(j)}u, \qquad b_1' z(1) = v, \qquad z(j-1)z(j) = p^{n-n(j)}u,$$

for $1 < j \leqq k$ and by the equations

$$b_j' \circ b_j'' = p^{n-m(j)}u, \qquad b_1' \circ z(1) = v, \qquad z(j-1) \circ z(j) = p^{n-n(j)}u,$$

for $1 < j \leqq k$, $b_2' \circ z(1) = p^{m-q}v$, where $q = \min(m(2), n(1)) \neq 0$. (Note that $1 \leqq k$, and that $k = 1$ is a possibility.)

The elements in $W(xy, G)$ are exactly the elements contained in the following two sets A and B: A consists of the elements of the form

$$p^m w(1)'b_1' + w(1)''b_1'' + \sum_{j=2}^{i} (w(j)'b_j' + w(j)''b_j'') + \sum_{j=2}^{k} w(j)z(j),$$

and B consists of the elements of the form

$$\begin{aligned} &w(1)''b_1'' + hz(1) && \text{if } k = 1, \\ &w(1)''b_1' + hz && \text{if } 1 < k, \end{aligned}$$

where

$$z = \sum_{j=0}^{i} p^{h(j)}z(2j+1),$$

$$h(j) = (n(2j+2) - n(2j+3)) + \cdots + (n(k-1) - n(k))$$

if $k = 2i+1$; and

$$z = \sum_{j=0}^{i-1} p^{h(j)}z(j),$$

$$h(j) = (n(2j+2) - n(2j+3)) + \cdots + (n(k-2) - n(k-1)) + n(k)$$

if $k = 2i$.

In order to show that $W(x \circ y, G)$ is essentially different from $W(xy, G)$ and that consequently the two proper multiplications xy and $x \circ y$ of G in V are not isomorphic, the following remark will suffice:

$$A' \leqq W(x \circ y, G) \leqq (A', B'),$$

where (A', B') may denote the set-theoretical join of the sets A' and B'. Here A' consists of the elements

$$w(1)'p^{m-q}b_1' + (w(2)'p^q - w(1)')b_2' + w(1)''b_1'' + w(2)''b_2'' + \sum_{j=3}^{i} (w(j)'b_j' + w(j)''b_j'') + \sum_{j=2}^{k} w(j)z(j),$$

and B' consists of the elements

$$w(1)'p^{m-q}b_1' + (w(2)'p^q - w(1)')b_2' + hz(1),$$

if $k=1$, and of the elements

$$w(1)'p^{m-q}b_1' + (w(2)'p^q - w(1)')b_2' + hz,$$

if $1<k$, where z has the same meaning as in the computation of $W(xy, G)$.

12. We prove the following statement:

(12.1) *If G is a direct sum of infinitely many infinite cyclic groups, and if the orders of the elements in $F(V)$ are not bounded, then there exists a proper multiplication of G in V.*

Proof. Clearly it is sufficient to prove the statement for countable groups G. Then there exists a countable basis B of G, and the elements of such a basis B may be denoted by $(i_1, \cdots, i_n)$ where n and i_j run over all the positive integers. There exists furthermore to every positive integer i an element $c(i)$ in V whose order exceeds i. Then a proper multiplication of G in V is characterized by the equations

$$(i_1, \cdots, i_n)(i_1, \cdots, i_n, i_{n+1}) = c(i_{n+1}).$$

(12.2) *If G is a direct sum of infinite cyclic groups, if $r(G, 0) = r(G)$ is not an odd finite number, and if V contains elements of infinite order, then there exists a proper multiplication of G in V. If $2<r(G, 0)$, then there exists a proper multiplication xy of G in V such that $(G<G; 2M(G, xy), xy) = 2G$ and a proper multiplication $x \circ y$ of G in V such that $2G<(G<G; 2M(G, x \circ y), x \circ y)$; and these two proper multiplications of G in V are not isomorphic.*

Proof. There exists a basis of G which consists of pairs b_v', b_v''. There exists furthermore an element u of infinite order in V, and a proper multiplication of G in V is characterized by the equations

$$b_v' b_v'' = u$$

for every v. Clearly

$$2G = (G < G; 2M(G, xy), xy).$$

If $2<r(G, 0)$, then a proper multiplication of G in V is characterized by the equations

$$b_1' \circ b_1' = u, \qquad b_v' \circ b_v'' = 2u$$

for every $v \neq 1$. Since $(G<G; 2M(G, x \circ y), x \circ y)$ is generated by the elements $2b_1'$, $2b_1''$, b_v', b_v'', for $v \neq 1$, it follows that $2G<(G<G; 2M(G, x \circ y), x \circ y)$. This completes the proof.

(12.3) *If G is a direct sum of three infinite cyclic groups, and if V is a direct sum of two infinite cyclic groups and one cyclic group (which may be 0, finite, or infinite), then there exists a proper multiplication xy of G in V such that $M(G, xy) = V$.*

Proof. There exists a basis b', b'', b''' of G, and there exists a basis u', u'', v of V such that u' and u'' are elements of infinite order. Then a proper multiplication of G in V is characterized by the equations

$$b'b'' = u', \qquad b''b''' = u'', \qquad b'''b' = v;$$

and clearly $V = M(G, xy)$.

(12.4) *If G is a direct sum of four infinite cyclic groups, and if V is a direct sum of an infinite cyclic group and one cyclic group, then there exists a proper multiplication xy of G in V such that $M(G, xy) = V$.*

Proof. There exists a basis $b_1, \cdots, b_4$ of G and a basis u, v of V, where u may be an element of infinite order. A proper multiplication xy of G in V such that $V = M(G, xy)$ is characterized by the equations $b_1b_2 = b_3b_4 = u$, $b_2b_3 = v$.

13. We prove the following statement:

(13.1) *If G is a direct sum of cyclic groups, if $G \neq F(G)$, if the orders of the elements in $F(G)$ are not bounded, and if V contains elements of order n whenever $F(G)$ contains elements of order n, then there exists a proper multiplication of G in V.*

Proof. $G = F(G) + U' + U''$, where U' is a direct sum of a finite number (not zero) of infinite cyclic groups and where U'' is either 0 or a direct sum of an infinity of infinite cyclic groups. Since the orders of the elements in $F(V)$ are not bounded, it follows from (12.1) that there exists a proper multiplication of U'' in V. Therefore it may be assumed without loss in generality that $r(G, 0)$ is finite.

If $r(G, 0)$ is finite, then G is a direct sum of $r(G, 0)$ groups $H(i)$ such that the orders of the elements in $F(H(i))$ are not bounded and such that $r(H(i), 0) = 1$. Thus it may be assumed without loss in generality that $r(G, 0) = 1$.

$F(G)$ is a direct sum of two groups G' and G'' such that G'' is a direct sum of two isomorphic groups and such that the orders in G' are not bounded and all the numbers $r(G', p^i)$ are finite. Since there exists by (11.1) a proper multiplication of G'' in V, it is sufficient to prove the existence of a proper multiplication of $G' + \bar{u}$ in V, where u is an element in G which generates $G \pmod{F(G)}$ and where $\bar{u}$ is the group generated by u.

If $F(G', p) \neq 0$, then there exists a basis $b_{1p}, b_{2p}, \cdots, b_{ip}, \cdots$ of $F(G', p)$

such that the order of b_{i-1p} is not greater than the order of b_{ip}. There exists, furthermore, in V an element v_{ip} of the same order as b_{ip}.

A proper multiplication of $G'+\bar{u}$ in V is characterized by the following equations

$$b_{i-1p}b_{ip} = v_{i-1p},$$
$$b_{ip}u = v_{ip}$$

for every i, if $F(G', p)$ is infinite, and

$$b_{i-1p}b_{ip} = v_{i-1p},\ b_{mp}u = v_{mp},$$

if $F(G', p)$ is finite and $m = r(F(G', p)_p)$.

(13.2) *If G is a direct sum of a finite number of cyclic groups, if $1 < r(G, 0)$, if V contains elements of the finite order n whenever $F(G)$ contains elements of order n, if V contains elements of infinite order, and if V contains two independent elements of infinite order, in case $r(G, 0)$ is odd, then the following propositions are true*:

(1) *There exists a proper multiplication xy of G in V with the properties*:

(i) *$(F(G, p) < F(G, p); 0, xy)$ is a cyclic group which is 0 if, and only if, $F(G, p)$ is a direct sum of an even number of isomorphic cyclic groups*;

(ii) *$M((F(G) < G; 0, xy), xy)$ does not contain non-zero elements of finite order.*

(2) *There exists a proper multiplication xy of G in V with the property*:

(iii) *$(F(G, p) < F(G, p); 0, xy)$ is a cyclic group, not zero, if $F(G, p)$ is a direct sum of an odd number of isomorphic cyclic groups, and it is a direct sum of two cyclic groups, not zero, otherwise, if $F(G, p) \neq 0$.*

(3) *If $F(G, p) \neq 0$, and if V contains two independent elements of order p, then there exists a proper multiplication xy of G in V such that $M(G, xy)$ contains two independent elements of order p.*

(4) *If V contains elements of finite order, and if $2 < r(G, 0)$, then there exists a proper multiplication xy of G in V such that $M((F(G) < G; 0, xy), xy)$ contains non-zero elements of finite order.*

Proof. There exists a basis $b_1, b_2, \cdots, b_h$ of G (mod $F(G)$) such that $1 < h = r(G, 0)$; and there exists a basis $b_{1p}, \cdots, b_{kp}$ of $F(G, p)$ such that $0 \leqq k = k(p)$ (that is, $k(p) = 0$, if $F(G, p) = 0$) and such that the orders $p^{n_{ip}}$ of b_{ip} satisfy the inequality $0 < n_{jp} \leqq n_{j-1p}$, for $0 < j - 1 < k$. V contains an element $v(p)$ of order $p^{n_{1p}}$ if $F(G, p) \neq 0$, an element u of infinite order, and, if $r(G, 0)$ is odd, an element v of infinite order which is independent of u.

A proper multiplication xy of G in V is characterized by the following equations:

$$b_1b_2 = b_3b_4 = \cdots = u;$$
$$b_{h-1}b_h = v$$

if, and only if, h is odd; and

$$b_2b_h = w,$$

if $2<h$, where w is an element of finite order in V which shall be kept indeterminate for the moment. If $F(G, p)\neq 0$, then

$$b_{1p}b_1 = v(p), \qquad b_{1p}b_2 = w(p),$$

where $w(p)$ is either 0 or an element of order p which is independent of $v(p)$, and

$$b_{j-1p}b_{jp} = p^{m_{jp}}v(p) \quad \text{with} \quad m_{jp} = n_{1p} - n_{jp}.$$

If $w=0$ and $w(p)\neq 0$, then xy meets the requirements of (3). If $w(p)=0$, then $(F(G)<G;\ 0,\ xy)$ is the direct sum of $(F(G)<F(G);\ 0,\ xy)$, the cyclic groups generated by the elements b_i for $1<i$, and the cyclic group generated by n_1b_1, where $n_1=\prod_p p^{n_{1p}}$. The requirements of (4) are therefore satisfied if $w\neq 0$; and condition (ii) is satisfied if $w=0$.

Suppose now that $w=w(p)=0$. Every element of $F(G, p)$ has the form $x=\sum_{j=1}^{k(p)}x_jb_{jp}$, consequently

$$xb_{1p} = -x_2p^{m_{2p}}v(p), \qquad xb_{jp} = (x_{j-1}p^{m_{jp}} - x_{j+1}p^{m_{j+1p}})v(p)$$

for $1<j<k(p)$, and

$$xb_{k(p)p} = x_{k(p)-1}p^{m_{kp}}v(p).$$

Thus x belongs to $(F(G, p)<F(G, p);\ 0,\ xy)$ if, and only if,

$$0 = x_2b_{2p} = x_4b_{4p} = \cdots, \quad x_{j-1} = x'_{j-1}p^{n_{jp}-n_{j+1p}},$$
$$x'_{j-1} \equiv x_{j+1} \pmod{p^{n_{j+1p}}}, \qquad x_{k(p)-1} \equiv 0 \pmod{p^{n_{kp}}}.$$

Consequently $(F(G, p)<F(G, p);\ 0,\ xy)$ is the cyclic group generated by

$$\sum_{j=0}^{i-1} p^{h_{jp}}b_{2(i-j)-1p}$$

with

$$h_{0p} = n_{2ip}, \cdots, h_{jp} = n_{2ip} + (n_{2i-2p} - n_{2i-1p}) + \cdots + (n_{2(i-j)p} - n_{2(i-j)+1p})$$

if $k(p)=2i$, and by

$$\sum_{j=0}^{i} p^{h_{jp}}b_{2(i-j)+1p}$$

with

$$h_{ip} = [n_{2(i-j)p} - n_{2(i-j)+1p}] + \cdots + [n_{2ip} - n_{2i+1p}]$$

if $k(p) = 2i+1$. It is now obvious that this multiplication xy satisfies (i).

In order to prove (2) we will consider the multiplication $x \circ y$ of G in V which is characterized by the equations

$$b_1 \circ b_2 = b_3 \circ b_4 = \cdots = u,$$
$$b_{h-1} \circ b_h = v,$$

if and only if h is odd; if $F(G, p) \neq 0$, then

$$b_{1p} \circ b_1 = v(p), \qquad b_{j-1p} \circ b_{jp} = p^{m_{jp}} v(p), \quad \text{for} \quad j < k(p),$$
$$b_{k(p)p} \circ b_2 = p^{m_{k(p)p}} v(p), \quad \text{if} \quad 1 < k(p).$$

If $k(p) = 1$, then $(F(G, p) < F(G, p);\ 0, x \circ y)$ is the cyclic group generated by b_{1p}. If $1 < k(p)$, then $(F(G, p) < F(G, p);\ 0, x \circ y)$ is the direct sum of the cyclic group, generated by $b_{k(p)p}$, and the group $W(p)$, consisting of all the elements w in the group B, generated by the b_{jp} with $j < k(p)$, which satisfy $wx = 0$ for every x in $F(G, p)$. But $W(p) = (B < B;\ 0, x \circ y)$, and in the group, generated by B and the b_j for $j < k(p)$, the multiplication is of the same type as considered in the first part of the proof. Thus $W(p)$ is a cyclic group, and $W(p) = 0$ if, and only if, B is a direct sum of an even number of isomorphic cyclic groups. This completes the proof.

14. **Proof of the existence theorem.** If G is a direct sum of cyclic groups, and if V is an abelian group, then it follows from the transformation of the existence theorem (§3) that the existence theorem is equivalent to the following proposition:

There exists a proper multiplication of G in V if, and only if, the conditions (i) *to* (vi) *of the existence theorem are satisfied.*

Suppose first that there exists a proper multiplication xy of G in V. Then (i) and (ii) are consequences of (5.2). Suppose now that the orders of the elements in $F(G)$ are bounded and that $r(G, 0)$ is a finite positive integer. Then $G = F(G) + U$, where $U \neq 0$ is a direct sum of a finite number of infinite cyclic groups. If m is the finite maximum order of the elements in $F(G)$, then $muv = 0$ for u in $F(G)$ and v in U, and it is now a consequence of (6.2) that xy induces a proper multiplication of U in V. Since U is generated by a finite number of elements, $M(U, xy)$ is also generated by a finite number of elements, and $F(M(U, xy))$ is therefore a finite group. Hence xy induces by (6.1) a proper multiplication of U in $M(U, xy)/F(M(U, xy)) = V^*$, and all the non-zero elements in V^* are of infinite order. Now it follows from (5.2) that $V^* \neq 0$, that is, that V contains elements of infinite order, and it follows from (5.1) that U contains at least two independent elements, that is, $1 < r(G, 0)$. If

furthermore $r(G, 0)$ is an odd finite number, then it follows from Corollary 7.2 that V^* is not a cyclic group, and consequently that V contains at least two independent elements of infinite order. Thus the necessity of the conditions (iii) and (iv) has been verified.

If $G=F(G)$, then $G=F(G, p)+F'(G, p)$, where $F'(G, p)$ is the direct sum of all the $F(G, q)$ for $q\neq p$, and $uv=0$ for u in $F(G, p)$ and v in $F'(G, p)$. Thus a multiplication xy of G in V is a proper multiplication if, and only if, xy induces a proper multiplication of every $F(G, p)$ in V.

Suppose now that xy is a proper multiplication of $F(G, p)$ in V and that the orders of the elements in $F(G, p)$ are bounded. Then (v) is a consequence of (5.2) and (5.1). Suppose now that $F(G, p)=F'+F''$, where $p^{i-1}F'=0$, whereas F'' is a direct sum of cyclic groups whose orders are multiples of p^i. Then xy induces by (6.2) a proper multiplication of F'' in V. Clearly $M(F'', xy)\leqq F(V, p)$. If $F(V, p)=V'+V''$, where $p^{i-1}V'=0$, V'' contains exactly one cyclic subgroup of order p^i, then it follows from (6.1) that xy induces a proper multiplication of F'' in $F(V, p)/V'\sim V''$. If F'' is a finite group, it follows from Corollary 7.2 that F'' is a direct sum of two isomorphic groups. This shows the necessity of condition (vi).

Assume now that the direct sum G of cyclic groups and the abelian group V satisfies the conditions (i) to (vi) of the existence theorem.

Case 1. $F(G)\neq G$. If the orders of the elements in $F(G)$ are not bounded, there exists, by (13.1) and (i), a proper multiplication of G in V. If the orders of elements in $F(G)$ are bounded, then $G=F'+F''+U$, where F' is a finite group, F'' is a direct sum of two isomorphic groups without elements of infinite order, and U is a direct sum of infinite cyclic groups. If V does not contain elements of infinite order, then U is, by (iii), a direct sum of an infinity of infinite cyclic groups, and the orders of the elements in $F(V)$ are not bounded, by (ii). There exists, by (12.1), a proper multiplication xy of U in V and there exists, by (11.1), a proper multiplication xy of F'' in V. Consequently there exists a proper multiplication of $F''+U$ in V. F' has, as a finite group, a basis of the form $b_1, \cdots, b_k$, where the order of b_{i-1} is a divisor of the order of b_i. Denote by v_i an element in V whose order equals the order of b_i and by $u_0, u_1, \cdots$ a basis of U. Then the proper multiplication of $F''+U$ in V is extended to a proper multiplication of G in V by the equations

$$b_{i-1}b_i = v_i, \qquad b_k u_0 = v_k, \qquad b_j u_v = 0, \quad \text{for} \quad v \neq 0,$$

$b_j f''=0$ for f'' in F''.

If the orders of the elements in $F(G)$ are bounded, then $G=F'+U'+W$, where W is a direct sum of two isomorphic groups, F' a finite group, and U' a direct sum of two or three infinite cyclic groups; and if $r(U', 0)=3$, then

$r(W, 0)$ is finite. If the orders of the elements in $F(V)$ are bounded, then V contains elements of infinite order, and if $r(U', 0) = 3$, then $r(G, 0)$ is an odd finite number and V contains, by (iv), two independent elements of infinite order. Thus there exists, by (13.2), a proper multiplication of $F' + U'$ in V; and by (11.1), (12.2) there exists a proper multiplication of W in V. Consequently there exists a proper multiplication of G in V.

Case 2. $F(G) = G$. Since $F(G)$ is the direct sum of the groups $F(G, p)$, it is sufficient to construct proper multiplications of $F(G, p)$ in V. If the orders of the elements in $F(G, p)$ are not bounded, there exists by (11.2) a proper multiplication of G in V. If the orders of the elements in $F(G, p)$ are bounded, either $F(G, p)$ is a finite group and the existence of a proper multiplication of $F(G, p)$ in V is a consequence of (11. 3), or $F(G, p)$ is a direct sum of an infinity of cyclic groups of order p^n and some cyclic groups of lower order and the existence of a proper multiplication of $F(G, p)$ in V is a consequence of (11.4), or finally $F(G, p)$ is a direct sum of a group F' and a group F'', where F' is a direct sum of a finite number of cyclic groups whose orders are multiples of p^n whereas F'' is a direct sum of an infinity of cyclic groups of order p^{n-1} and cyclic groups of lower order. In this last case there exists a proper multiplication of $F(G, p)$ in V, since there exists by (11.3) a proper multiplication of F' in V and by (11.4) a proper multiplication of F'' in V.

15. **Theorem 15.1.** We prove the following theorem:

THEOREM 15.1. *Assume that V is an abelian group, and that G is a direct sum of a finite number of cyclic groups. Then there exists one and, apart from isomorphic ones, only one proper multiplication of G in V if the following conditions are satisfied:*

(A) *If $r(G, 0) \neq 0$, then $1 < r(G, 0) < 4$.*

(B) *If $r(G, 0) = 3$, then $F(G) = F(V) = 0$ and $2 = r(V, 0)$ $(= r(V))$.*

(C) *If $r(G, 0) = 2$ and $F(G, p) \neq 0$, then $F(G, p)$ is a direct sum of an odd number of isomorphic cyclic groups, and V contains elements of infinite order and exactly one cyclic subgroup of order p^k, where p^k is the order of the cyclic direct summands of $F(G, p)$.*

(D) *If $G = F(G)$ (that is, if G is a finite group), if $F(G, p) \neq 0$ and if $2 < r(V_p)$, then $F(G, p)$ is a direct sum of two isomorphic cyclic groups.*

(E) *If $G = F(G)$, $F(G, p) \neq 0$, and $2 = r(V_p)$, then $F(G, p)$ is a direct sum of two cyclic groups of order p^n and one cyclic group of order p^m with $0 \leqq m \leqq n$, $0 < n$, and V contains two independent elements of order p^m.*

(F) *If $G = F(G)$, $F(G, p) \neq 0$, and $1 = r(V_p)$, then $F(G, p)$ is a direct sum of two isomorphic groups.*

(G) *If $F(G)$ contains elements of order n, then V contains elements of order n.*

Proof. If the conditions (A) to (G) are satisfied, then it follows from the existence theorem and its "transformation" that there exists a proper multiplication of G in V. That any two proper multiplications of G in V are isomorphic, if the conditions (A) to (G) are satisfied, is a consequence of the following propositions:

(1) Lemma 8.1 if $r(G, 0)=3$.

(2) Corollary 7.4 if $r(G, 0)=2$.

(3) Corollary 7.2 if $G=F(G)$ is a direct sum of two isomorphic groups and $r(V_p)=1$ for every prime number p such that elements of order p are contained in G.

(4) Lemma 8.1 if $G=F(G)$ and $F(G, p)$ is a direct sum of at most three cyclic groups and $r(V_p)=2$, provided $F(G, p)\neq 0$.

(5) (5.1) and (5.2) if $G=F(G)$ and if $F(G, p)$ is a direct sum of two isomorphic cyclic groups.

Suppose now that there exists one, and apart from isomorphic multiplications, only one proper multiplication of G in V. Assume first that $F(G)\neq G$. Then $1<r(G, 0)$ since $F(G)$ and $r(G, 0)$ are finite, as follows from condition (iii) of the existence theorem. Since furthermore $G=F(G)+U+U'$, where U is a direct sum of an even number of infinite cyclic groups and U' is either 0 or an infinite cyclic group, and since every proper multiplication of U in an infinite cyclic subgroup of V may be extended to a proper multiplication of G in V, it follows from (12.2) that $r(G, 0)\leqq 3$. If $r(G, 0)=3$, then it follows from (13.2), (1) and (4), that $F(V)=0$, and it is a consequence of condition (i) of the existence theorem that $F(G)=0$. Now condition (B) is a consequence of (12.3). If $r(G, 0)=2$, then it follows from (13.2), (1) and (2), that $F(G, p)$ is either 0 or a direct sum of an odd number of isomorphic cyclic groups and that $r(V_p)=1$ is a consequence of (13.2), (3). Thus conditions (A) to (C) are proved to be necessary.

Assume secondly that $G=F(G)$ is a finite group. Then the conditions (D), (E), (F) are consequences of (11.4) and of the conditions (i), (v), and (vi) of the existence theorem which imply also condition (G).

16. **Proof of the uniqueness theorem.** If G is a direct sum of a finite number of cyclic groups, if V is an abelian group, and if G and V satisfy the conditions (1) to (8) of the uniqueness theorem, then it is a consequence of Theorem 15.1 that any two proper multiplications of G in V are isomorphic and that proper multiplications of G in V exist. If G contains elements of infinite order, then V is a direct sum of groups of type* p^∞ and of groups of the type of the additive group of all the rational numbers, and if $G=F(G)$, but

* Groups of type p^∞ contain only elements whose orders are powers of p, and contain, for every positive integer i, exactly one cyclic subgroup of order p^i.

$F(G, p) \neq 0$, then $F(V, p)$ is a direct summand of V which is itself a direct sum of groups of type p^∞.* Thus the isomorphisms of relevant subgroups of finite rank of V may be induced by automorphisms of V. Since finally all the functions $P(n, x)$ appearing in the transformation of the uniqueness theorem satisfy $P(n, x) = 0$ (because $V = nV$ for every relevant n), it follows from the transformation of the uniqueness theorem that the conditions (1) to (8) are sufficient.

In order to prove the necessity of the conditions (1) to (8) it is sufficient to prove the necessity of the conditions (1) and (2), as may be inferred from Theorem 15.1 and the transformation of the uniqueness theorem. V contains a subgroup of the form $F(V) + U$, where U is a direct sum of $r(G, 0)$ infinite cyclic groups, and where U may contain any preassigned element of infinite order as a basis element. If $F(G) \neq G$, then there exist proper multiplications xy of G in $F(V) + U$ such that a given basis element of U appears in $M(G, xy)$. If $V \neq nV$ for some positive n, then it is always possible to find a proper multiplication xy of G in V such that all elements of infinite order in $M(G, xy)$ are elements in nV and to find another one where this is not the case. Thus condition (2) is also necessary.

In order to prove (1) we will have to consider the functions $P(n, x)$. For every proper multiplication xy of G in V there exist admissible functions $P(n, x)$ which have, on a basis of $F(G)$, preassigned values. If $P(n, x) = 0$ for every x of a basis of $F(G)$, then $P(n, x) = 0$ for every odd n and every x in G_n, and the $P(2^i, x)$ are elements in $M(G, xy)_2$. Thus $V = pV$ for every odd prime number such that $F(G, p) \neq 0$; and if V contains elements of order 4 and if $F(G, 2) \neq 0$, then $V = 2V$. By condition (2) and condition (i) of the existence theorem only the following case has to be discussed in order to complete the proof of the necessity of condition (1): *$F(G, 2)$ is a direct sum of groups of order 2 and so is $F(V, 2)$.*

It follows from (6) to (8) that only the following cases are possible:

I. $r(F(G, 2)) = 3$ and $r(F(V, 2)) = 2$.
II. $2 < r(F(G, 2)) = 2i$ and $r(F(V, 2)) = 1$.
III. $2 = r(F(G, 2))$.

Case I. There exists, by Lemma 8.1, for the given proper multiplication xy of G in V a basis b', b'', b of $F(G, 2)$ such that bb' and $b'b''$ form a basis of $F(G, 2)$ and such that $bb'' = 0$. If $P(2, b) = P(2, b') = P(2, p'') = 0$, then

$$P(2, b + b') = bb', \qquad P(2, b' + b'') = b'b'', \qquad P(2, b'' + b) = 0,$$
$$P(2, b + b' + b'') = bb' + b'b''.$$

* Cf. R. Baer, Annals of Mathematics, vol. 37 (1936), pp. 766–781.

If on the other hand $P'(2, b) = P'(2, b') = P'(2, b'') = bb'$, then

$$P'(2, b + b') = bb', \qquad P'(2, b' + b'') = b'b'', \qquad P'(2, b'' + b) = 0,$$
$$P'(2, b + b' + b'') = b'b'';$$

and these two sets of admissible functions are clearly essentially different, unless $V = 2V$.

Case II. There exists, by Corollary 7.2, a basis b_i', b_i'' of $F(G, 2)$ such that $b_1' b_1'' = \cdots = b_k' b_k'' = c$ and such that all the other products of basis elements are 0. The admissible function $P(2, x)$, characterized by $P(2, b_i') = P(2, b_i'') = 0$, has the property that $P(2, x) \neq 0$ for exactly 2^{k-1} elements x in $F(G, 2)$. The admissible functions $P'(2, x)$, on the other hand, which are characterized by

$$P'(2, b_1') = P'(2, b_1'') = c, \qquad P'(2, b_j') = P'(2, b_j'') = 0,$$

for $1 < j$, have the property that $P'(2, x) \neq 0$ for exactly $2^{k-1}(3 \cdot 2^{k-1} - 1)$ (for 3, if $k = 1$) elements x in $F(G, 2)$. This completes the proof of the necessity of condition (1) since this treatment of Case II covers Case III too.

THE UNIVERSITY OF ILLINOIS,
URBANA, ILL.

DEFINITELY SELF-ADJOINT BOUNDARY VALUE PROBLEMS*

BY

GILBERT A. BLISS

1. **Introduction.** The boundary value problem to be considered in this paper is that of finding a constant λ and a set of functions $y_i(x)$ $(a \leqq x \leqq b;\ i=1, \cdots, n)$ satisfying differential equations and boundary conditions of the form

$$(1.1) \qquad y_i' = [A_{i\alpha}(x) + \lambda B_{i\alpha}(x)]y, \qquad M_{i\alpha}y_\alpha(a) + N_{i\alpha}y_\alpha(b) = 0,$$

in which the matrix $\|M_{i\alpha}N_{i\alpha}\|$ is a matrix of real constants of rank n. Repeated subscripts indicate summation, as in tensor analysis, and it will be understood that all subscripts have the range $1, \cdots, n$ unless otherwise explicitly specified. The system

$$(1.2) \qquad z_i' = -z_\alpha[A_{\alpha i} + \lambda B_{\alpha i}], \qquad z_\alpha(a)P_{\alpha i} + z_\alpha(b)Q_{\alpha i} = 0$$

is by definition *adjoint* to (1.1) if the matrix of constants $\|P_{\alpha i}Q_{\alpha i}\|$ satisfies the conditions

$$M_{i\alpha}P_{\alpha k} - N_{i\alpha}Q_{\alpha k} = 0 \qquad (i, k = 1, \cdots, n).$$

The system which is given in (1.1) is said to be *self-adjoint* provided that it is equivalent to its adjoint system (1.2) by a non-singular transformation $z_i = T_{i\alpha}(x)y_\alpha$.

This definition of self-adjoint boundary value problems and a further definition of so-called definite self-adjointness were given by the author in a paper published in 1926† which will be designated in the text below by the Roman numeral I. In that paper it was stated that the boundary value problems arising from the calculus of variations are all definitely self-adjoint. This statement is true for non-singular problems of the calculus of variations without side conditions, the only ones whose boundary value problems had been studied up to that time so far as is known to the writer. It is not true, however, for problems of the calculus of variations such as those of Mayer, Lagrange, and Bolza whose boundary value problems are self-adjoint but not definitely self-adjoint according to the definition given in I. One of the earliest

* Presented to the Society, April 11, 1936; received by the editors November 3, 1937.

† Bliss, *A boundary value problem for a system of ordinary differential equations of the first order*, these Transactions, vol. 28 (1926), pp. 561–584.

formulations of a case of this more complicated kind was that of Cope for the problem of Mayer with variable end points.*

In the following pages a modification of the earlier definition of definite self-adjointness will be given which seems to be applicable to all of the boundary value problems so far studied arising from problems of the calculus of variations involving simple integrals. The new definition involves a property analogous to the normality of a minimizing arc for a problem of Bolza, and is weaker than the older definition in the sense that it imposes fewer restrictions. It will be shown, however, that for a definitely self-adjoint boundary value problem as here defined most of the properties deduced in the paper I cited above are still valid. For example, the characteristic numbers are all real and have indices equal to their multiplicities, and the expansion theorems proved in the paper I also hold. It is not possible to show that the number of characteristic numbers is always infinite. Examples will be cited showing that this is in fact not the case. When the set of characteristic numbers is finite the class of functions for which the expansion theorems hold is of course severely limited. The boundary value problems arising from the calculus of variations are a special type of definitely self-adjoint problems which have an infinity of characteristic numbers, as has been shown by several writers.†

In the paragraphs below frequent use is made of the results and proofs of the paper I to which reference has been made above.

2. **The definition of definite self-adjointness and its first consequences.** It is understood that $A_{ik}(x), B_{ik}(x)$ are real, single-valued and continuous on $a \leqq x \leqq b$. The definition fundamental for this paper is then the following:

DEFINITION. A boundary value problem (1.1) is said to be *definitely self-adjoint* if it is self-adjoint and has the further properties:

(1) the matrix of functions $S_{ik}(x) = T_{\alpha i}(x)B_{\alpha k}(x)$ is symmetric at each value x on the interval ab;

(2) the quadratic form $S_{\alpha\beta}(x)\xi_\alpha\xi_\beta$ is non-negative at each value x on ab;

(3) the set $y_i(x) \equiv 0$ is the only set of functions which satisfies on ab the conditions

$$(2.1) \qquad y_i' = A_{i\alpha}y_\alpha, \qquad M_{i\alpha}y_\alpha(a) + N_{i\alpha}y_\alpha(b) = 0, \qquad S_{\alpha\beta}y_\alpha y_\beta = 0.$$

* Cope, *An analogue of Jacobi's condition for the problem of Mayer with variable end-points*, Dissertation, University of Chicago, 1927. For a synopsis see *Abstracts of Theses*, The University of Chicago, vol. 6 (1927–1928), pp. 15–21. See also American Journal of Mathematics, vol. 59 (1937), pp. 655–672.

† Hu, *The problem of Bolza and its accessory boundary value problem*, *Contributions to the Calculus of Variations* 1931–1932, The University of Chicago Press, p. 400; Morse, *Sufficient conditions in the problem of Lagrange with variable end conditions*, American Journal of Mathematics, vol. 53 (1931), pp. 517–546, especially §16.

The property (3) is analogous to normality in the calculus of variations, as will be shown in a later section. Since the quadratic form with matrix S_{ik} is non-negative, it follows that every set of values y_i which satisfy the equation $S_{\alpha\beta}y_\alpha y_\beta=0$ must also satisfy $S_{i\alpha}y_\alpha=0$ and consequently the equations $B_{i\alpha}y_\alpha=0$, since the determinant $|T_{ik}|$ is different from zero. In the conditions (2.1) we can therefore replace the last equation by $B_{i\alpha}y_\alpha=0$ if desirable.

Let $Y_{ik}(x, \lambda)$ be the elements of a matrix whose columns are n linearly independent solutions of the differential equations in (1.1), and let $s_i(y)$ represent the first member of the second equation (1.1). Then the *characteristic numbers* of the boundary value problem are the roots of the determinant given by

$$D(\lambda) = |\, s_i[Y_k(x, \lambda)]\,|$$

in which the symbol $s_i[Y_k(x, \lambda)]$ represents the value of $s_i(y)$ formed for the kth column of the matrix of elements Y_{ik}. The *index* of a root λ_0 of $D(\lambda)$ is by definition the number r when $n-r$ is the rank of $D(\lambda_0)$, and the *multiplicity* of λ_0 is its multiplicity as a root of $D(\lambda)$.

THEOREM 2.1. *For a definitely self-adjoint boundary value problem every root of the determinant $D(\lambda)$ is real, and the independent characteristic solutions of the boundary value problem corresponding to such a root may be chosen real.*

For suppose that $y_i=y_{i1}+(-1)^{1/2}y_{i2}$ were a solution of the boundary value problem, not identically zero and corresponding to an imaginary root $\lambda=\lambda_1+(-1)^{1/2}\lambda_2$ of $D(\lambda)$. Then the conjugate imaginary set $\bar{y}_i=y_{i1}-(-1)^{1/2}y_{i2}$ would be a solution corresponding to the root $\bar{\lambda}=\lambda_1-(-1)^{1/2}\lambda_2$. According to I, Theorem 8, we would have

$$S_{\alpha\beta}y_\alpha\bar{y}_\beta = S_{\alpha\beta}y_{\alpha 1}y_{\beta 1} + S_{\alpha\beta}y_{\alpha 2}y_{\beta 2} = 0.$$

This would imply a contradiction since by a remark made above the equations $B_{i\alpha}y_{\alpha 1}=B_{i\alpha}y_{\alpha 2}=0$ would be consequences of the last equation, and one verifies readily by substitution in (1.1) that the functions $y_{i1}(x)$ and $y_{i2}(x)$ would satisfy the equations (2.1) and hence be identically zero.

THEOREM 2.2. *For a definitely self-adjoint boundary value problem the index of every root of $D(\lambda)$ is equal to its multiplicity.*

The proof is identical with that of I, Theorem 10, down to the last equation on page 572 which would again imply $B_{i\alpha}y_{\alpha 1}\equiv 0$ and $y_{i1}\equiv 0$, as in the paragraph above preceding Theorem 2.2, and this would be a contradiction since the functions y_{i1} in the proof are not identically zero.

For the new definition of definite self-adjointness the Theorem 11 of the paper I will be replaced by the following theorem which is analogous to a

theorem of Hu* for boundary value problems of the calculus of variations:

THEOREM 2.3. *If for a set of functions $f_i(x)$ continuous on the interval ab the condition*

$$\int_a^b S_{\alpha\beta} y_\alpha f_\beta dx = 0 \tag{2.2}$$

is satisfied by every solution $y_i(x)$ of a definitely self-adjoint boundary value problem, then it is also satisfied by every set of functions $y_i(x)$ satisfying the following equations

$$y_i' = A_{i\alpha} y_\alpha + B_{i\alpha} g_\alpha, \qquad s_i(y) = 0 \tag{2.3}$$

with functions $g_i(x)$ continuous on the interval ab.

The condition (2.2) for all solutions $y_i(x)$ of the boundary value problem implies, as in the proof of I, Theorem 11, that the non-homogeneous system

$$y_i' = (A_{i\alpha} + \lambda B_{i\alpha}) y_\alpha + B_{i\alpha} f_\alpha, \qquad s_i(y) = 0$$

has a solution $y_i(x, \lambda)$ expressible by power series

$$y_i(x, \lambda) = u_{i0}(x) + u_{i1}(x)\lambda + u_{i2}(x)\lambda^2 + \cdots$$

whose coefficients $u_{i\mu}(x)$ $(\mu = 0, 1, 2, \cdots)$ have continuous derivatives on the interval ab, and which converge uniformly for values x, λ satisfying conditions of the form $a \leqq x \leqq b$, $|\lambda| \leqq \rho$. From the proof of I, Theorem 11, it follows that

$$u_{i0}' = A_{i\alpha} u_{\alpha 0} + B_{i\alpha} f_\alpha, \qquad s_i(u_0) = 0, \tag{2.4}$$

and also that

$$W_0 = \int_a^b S_{\alpha\beta} u_{\alpha 0} u_{\beta 0} dx = 0.$$

From the last equation and the properties (1)–(3) we deduce the identities $B_{i\alpha} u_{\alpha 0} \equiv 0$. Consider now a set of functions $y_i(x)$ which satisfy equations of the form (2.3). From (2.3) and the equations (19), (20) of I it follows that the corresponding functions $z_i = T_{i\alpha} y_\alpha$ satisfy the equations

$$z_i' = -z_\alpha A_{\alpha i} - B_{\alpha i} T_{\alpha\beta} g_\beta, \qquad t_i(z) = 0, \tag{2.5}$$

where $t_i(z)$ is a symbol for the first member of the second equation (1.2). From (2.4) and (2.5) and the identities $B_{i\alpha} u_{\alpha 0} \equiv 0$, we have

$$z_\alpha u_{\alpha 0}' + z_\alpha' u_{\alpha 0} = S_{\alpha\beta} y_\alpha f_\beta,$$

* Hu, loc. cit., Theorem 7.3, p. 396.

and hence with the help of equation (7) of I we find that every set of functions y_i which satisfy the equations (2.3) will also satisfy (2.2), as was to be demonstrated.

COROLLARY 2.1. *If the determinant* $|B_{ik}(x)|$ *is different from zero on the interval ab, then* $f_i \equiv 0$ *is the only set of functions which satisfy the condition* (2.2) *with every solution* $y_i(x)$ *of a definitely self-adjoint boundary value problem.*

This follows from the equations (2.4) and the identities $u_{\alpha 0} \equiv 0$ which are consequences of the identities $B_{i\alpha} u_{\alpha 0} \equiv 0$ when the determinant $|B_{ik}|$ is nowhere zero.

COROLLARY 2.2. *If the functions* f_i *satisfy equations of the form*

$$f_i' = A_{i\alpha} f_\alpha + B_{i\alpha} g_\alpha, \qquad s_i(f) = 0$$

with functions $g_i(x)$ *continuous on the interval ab, and if they satisfy the condition* (2.2) *with every solution of a definitely self-adjoint boundary value problem, then they also satisfy the identities* $B_{i\alpha} f_\alpha \equiv 0$.

This follows readily from Theorem 2.3 when we note that the functions $y_i = f_i$ satisfy equations of the form (2.3), and therefore from (2.2) that

$$\int_a^b S_{\alpha\beta} f_\alpha f_\beta dx = 0.$$

By reasoning similar to that used a number of times above it follows then that $B_{i\alpha} f_\alpha \equiv 0$.

3. **The expansion theorems.** Since the roots of the power series $D(\lambda)$ form a finite or infinite denumerable set, and since the number of linearly independent solutions $y_i(x)$ of the boundary value problem associated with each root is equal to the multiplicity of the root, it follows that the solutions and their corresponding characteristic numbers can be enumerated and denoted by symbols $y_{i\nu}(x)$, λ_ν $(\nu = 1, 2, \cdots)$. Furthermore these solutions can be normed and orthogonalized by well known processes so that

$$\int_a^b S_{\alpha\beta} y_{\alpha\mu} y_{\beta\nu} dx = \delta_{\mu\nu}, \tag{3.1}$$

where $\delta_{\mu\mu} = 1$, $\delta_{\mu\nu} = 0$ if $\mu \neq \nu$. For an arbitrary set of functions $f_i(x)$ continuous on the interval ab the constants c_ν may be defined by the equations

$$c_\nu = \int_a^b S_{\alpha\beta}(\xi) y_{\alpha\nu}(\xi) f_\beta(\xi) d\xi. \tag{3.2}$$

The fundamental Theorem 13 of the paper I needs a different proof with the

new definition of definite self-adjointness. It may be written in the following form:

THEOREM 3.1. *For every set of functions $f_i(x)$ satisfying equations of the form*

$$f_i' = A_{i\alpha}f_\alpha + B_{i\alpha}g_\alpha, \qquad s_i(f) = 0 \tag{3.3}$$

with functions $g_i(x)$ continuous on the interval ab, the series

$$\phi_i = \sum_\nu c_\nu y_{i\nu}(x) \tag{3.4}$$

converge uniformly on the interval ab, and $B_{i\alpha}(f_\alpha - \phi_\alpha) \equiv 0$.

The sums ϕ_i may contain only a finite number of terms if the set of characteristic numbers λ_ν is finite. But the uniform convergence of these series can in every case be proved as in I, §6. To prove the identities $B_{i\alpha}(f_\alpha - \phi_\alpha) \equiv 0$ we note first that for every ν

$$\int_a^b S_{\alpha\beta}(f_\alpha - \phi_\alpha)y_{\beta\nu}dx = 0 \tag{3.5}$$

because of the equations (3.1) and (3.2). From Theorem 2.3 it follows therefore that

$$\int_a^b S_{\alpha\beta}(f_\alpha - \phi_\alpha)f_\beta dx = 0 \tag{3.6}$$

since the functions f_i by hypothesis satisfy the equations (3.3) which are of the form (2.3). Furthermore

$$\int_a^b S_{\alpha\beta}(f_\alpha - \phi_\alpha)\phi_\beta dx = 0 \tag{3.7}$$

because of the form of the functions (3.4) and the relations (3.5). By subtracting (3.7) from (3.6) we find that

$$\int_a^b S_{\alpha\beta}(f_\alpha - \phi_\alpha)(f_\beta - \phi_\beta)dx = 0;$$

hence, by the usual argument, we obtain the desired identities.

COROLLARY 3.1. *If the determinant $|B_{ik}(x)|$ is nowhere zero on the interval ab, then for every set of functions $f_i(x)$ having continuous derivatives on that interval and satisfying the boundary conditions $s_i(f) = 0$ the sums* (3.4) *converge uniformly and are equal to the functions f_i on the interval ab.*

The corollary is identical with Corollary 1 of paper I and is proved in the same way.

COROLLARY 3.2. *If the functions $f_i(x)$ satisfy equations of the form* (3.3), *and if furthermore the functions $g_i(x)$ in those equations are solutions of a similar system*

$$g_i' = A_{i\alpha}g_\alpha + B_{i\alpha}h_\alpha, \qquad s_i(g) = 0$$

with functions $h_i(x)$ continuous on the interval ab, then the series (3.4) *converge uniformly and are equal to the functions $f_i(x)$.*

This is Corollary 2 of I, page 576, but its proof needs emendation for the new definition of definite self-adjointness. We use the notations

$$d_\nu = \int_a^b S_{\alpha\beta}(\xi)y_{\alpha\nu}(\xi)g_\beta(\xi)d\xi, \qquad \psi_i = \sum_\nu d_\nu y_{i\nu}.$$

The equation

$$\phi_i' = A_{i\alpha}\phi_\alpha + B_{i\alpha}\psi_\alpha \tag{3.8}$$

is equation (39) of I and is proved in the same way. From (3.3), (3.8), and the fact that $s_i(y_\nu)=0$ it follows that

$$f_i' - \phi_i' = A_{i\alpha}(f_\alpha - \phi_\alpha) + B_{i\alpha}(g_\alpha - \psi_\alpha), \qquad s_i(f-\phi) = 0. \tag{3.9}$$

The last term in the first equation (3.9) vanishes identically since the equations $B_{i\alpha}(g_\alpha-\psi_\alpha)\equiv 0$ are consequences of Theorem 3.1 applied to the functions g_i in place of the f_i. The similar identities $B_{i\alpha}(f_\alpha-\phi_\alpha)\equiv 0$ for the functions f_i, from Theorem 3.1, imply that $(f_\alpha-\phi_\alpha)S_{\alpha\beta}(f_\beta-\phi_\beta)\equiv 0$ and hence from equations (3.9) and the property (3) in the definition of definite self-adjointness that $f_i-\phi_i\equiv 0$.

4. **The boundary value problem associated with a problem of Bolza.** The second variation of the problem of Bolza may be taken in the form

$$J_2(\xi, \eta) = 2\gamma[\xi_1, \eta(x_1), \xi_2, \eta(x_2)] + \int_{x_1}^{x_2} 2\omega(x, \eta, \eta')dx$$

in which 2γ is a homogeneous quadratic form in its $2n+2$ arguments ξ_1, $\eta_i(x_1)$, ξ_2, $\eta_i(x_2)$ $(i=1, \cdots, n)$, and 2ω is a homogeneous quadratic form in the $2n$ variables $\eta_i(x)$, $\eta_i'(x)$ with coefficients functions of x.

An accessory minimum problem associated with this second variation is that of finding in a class of sets ξ_1, ξ_2, $\eta_i(x)$, satisfying conditions of the form

$$\Phi_\beta(x, \eta, \eta') = 0 \qquad (\beta = 1, \cdots, m < n),$$

$$\Psi_\mu[\xi_1, \eta(x_1), \xi_2, \eta(x_2)] = 0 \qquad (\mu = 1, \cdots, p \leqq 2n+2),$$

$$\int_{x_1}^{x_2} \eta_i\eta_i dx = 1,$$

one which minimizes $J_2(\xi, \eta)$. The functions Φ_β are homogeneous and linear in the $2n$ variables η_i, η_i' with coefficients functions of x, and the functions Ψ_μ are p homogeneous linear independent expressions with constant coefficients in their $2n+2$ arguments.

The differential equations and end conditions for a minimizing set $\xi_1, \xi_2, \eta_i(x)$, for the accessory problem may be expressed with the help of the notations

$$\Omega = \omega + \mu_\beta \Phi_\beta, \qquad \zeta_i = \Omega_{\eta_i'}$$

as follows:*

$$d\Omega_{\eta_i'}/dx = \Omega_{\eta_i} - \lambda\eta_i, \qquad \Phi_\beta = 0, \tag{4.1}$$

$$\begin{aligned} \gamma_1 + \epsilon_\mu \Psi_{\mu,1} &= 0, \\ -\zeta_{i1} + \gamma_{i1} + \epsilon_\mu \Psi_{\mu,i1} &= 0, \\ \gamma_2 + \epsilon_\mu \Psi_{\mu,2} &= 0, \\ \zeta_{i2} + \gamma_{i2} + \epsilon_\mu \Psi_{\mu,i2} &= 0, \\ \Psi_\mu &= 0. \end{aligned} \tag{4.2}$$

The notations η_{is}, ζ_{is} $(s=1, 2)$ represent the values $\eta_i(x_s)$, $\zeta_i(x_s)$ $(s=1, 2)$. The subscripts 1, 2, $i1$, $i2$ attached to γ and Ψ_μ indicate partial derivatives with respect to $\xi_1, \xi_2, \eta_{i1}, \eta_{i2}$, respectively. It is to be shown that the equations (4.1) and (4.2) are equivalent to a boundary value problem of the type studied in the preceding sections.

The accessory minimum problem is said to be *non-singular* if the determinant of coefficients of the variables η_i, μ_β in the first members of the equations

$$\begin{aligned} \Omega_{\eta_i'}(x, \eta, \eta', \mu) &= \zeta_i, \\ \Phi_\beta(x, \eta, \eta') &= 0 \end{aligned} \tag{4.3}$$

is everywhere different from zero on the interval x_1x_2. It is said to satisfy the *non-tangency condition* if the equations $\Psi_\mu = 0$ have no non-vanishing solution $\xi_1, \xi_2, \eta_i(x_1), \eta_i(x_2)$ with $\eta_i(x_1) = \eta_i(x_2) = 0$, or, in other words, if the matrix of coefficients of ξ_1 and ξ_2 in the functions Ψ_μ is of rank 2. Finally the accessory problem is said to be *normal* if the only solution $\xi_1, \xi_2, \eta_i(x), \mu_\beta(x), \epsilon_\mu$ of equations (4.1) and (4.2) with $\eta_i(x) \equiv 0$ on the interval x_1x_2 is the one whose elements all vanish identically.†

* Bliss, *The problem of Bolza in the calculus of variations*, mimeographed lecture notes, The University of Chicago, 1935, p. 73, equations (14.1), and p. 76, equation (14.8).

† These definitions are customary ones. See, for example, Bliss, *The problem of Bolza in the calculus of variations*, loc. cit., pp. 34, 82, and §§9, 10. In the two sections cited the notion of normality is analyzed in considerable detail.

The differential equations (4.1) can be expressed in terms of the so-called canonical variables x, η_i, ζ_i related to the variables x, η_i, η_i', μ_β by means of the equations (4.3). If the accessory minimum problem is non-singular, these equations have solutions

$$\eta_i' = \Pi_i(x, \eta, \zeta), \qquad \mu_\beta = \mathrm{M}_\beta(x, \eta, \zeta)$$

which are linear in the variables η_i, ζ_i. In terms of the homogeneous quadratic form in η_i, ζ_i defined by the equation

$$\begin{aligned} 2\mathfrak{H} &= 2\zeta_i\Pi_i - 2\Omega(x, \eta, \Pi, \mathrm{M}) \\ &= U_{ij}(x)\eta_i\eta_j + 2V_{ij}(x)\eta_i\zeta_j + W_{ij}(x)\zeta_i\zeta_j \end{aligned}$$

the differential equations (4.1) take the well known canonical form

$$\begin{aligned} d\eta_i/dx &= \mathfrak{H}_{\zeta_i} = V_{ji}\eta_j + W_{ij}\zeta_j, \\ d\zeta_i/dx &= -\mathfrak{H}_{\eta_i} - \lambda\eta_i = -U_{ij}\eta_j - V_{ij}\zeta_j - \lambda\eta_i. \end{aligned} \tag{4.4}$$

The matrices of elements U_{ij} and W_{ij} are, of course, symmetric.

The end conditions (4.2) can also be transformed into a more convenient form. Let

$$\xi_1, \quad \xi_2, \quad \eta_{i1}, \quad \eta_{i2}, \quad \zeta_{i1}, \quad \zeta_{i2}, \quad \epsilon_\mu, \tag{4.5}$$

$$\alpha_1, \quad \alpha_2, \quad a_{i1}, \quad a_{i2}, \quad b_{i1}, \quad b_{i2}, \quad \theta_\mu \tag{4.6}$$

be two sets satisfying those conditions. If we multiply the first four equations in (4.2), respectively, by α_1, a_{i1}, α_2, a_{i2} and add, and then subtract the similar sum with the two solutions interchanged, it follows from the fifth equation (4.2) and well known properties of quadratic forms, that

$$b_{i1}\eta_{i1} - a_{i1}\zeta_{i1} - b_{i2}\eta_{i2} + a_{i2}\zeta_{i2} = 0. \tag{4.7}$$

Consider now $2n$ linearly independent solutions of equations (4.2)

$$\begin{matrix} \xi_{i,1}, & \xi_{i,2}, & \eta_{i,k1}, & \eta_{i,k2}, & \zeta_{i,k1}, & \zeta_{i,k2}, & \epsilon_{i,\mu}, \\ \alpha_{i,1}, & \alpha_{i,2}, & a_{i,k1}, & a_{i,k2}, & b_{i,k1}, & b_{i,k2}, & \theta_{i,\mu} \end{matrix} \tag{4.8}$$

for $i=1, \cdots, n$. If the accessory minimum problem satisfies the non-tangency condition, the elements η, ζ, a, b in the sets (4.8) form a $2n\times 4n$-dimensional matrix which is of rank $2n$. Otherwise there would be a solution (4.5) of equations (4.2) with elements η, ζ all zero, formed by taking a linear combination of the $2n$ solutions (4.8) with constant coefficients not all zero. The elements ξ_1, ξ_2 of this solution would also vanish, on account of the fifth of equations (4.2) and the non-tangency condition. The elements ϵ_μ would then also vanish because of the first four of equations (4.2) and the independence

of the functions Ψ_μ. This would contradict the independence of the $2n$ solutions (4.8). It is evident then that the $2n$ equations

$$\begin{aligned} &\zeta_{i,k1}\eta_{k1} - \eta_{i,k1}\zeta_{k1} - \zeta_{i,k2}\eta_{k2} + \eta_{i,k2}\zeta_{k2} = 0, \\ &b_{i,k1}\eta_{k1} - a_{i,k1}\zeta_{k1} - b_{i,k2}\eta_{k2} + a_{i,k2}\zeta_{k2} = 0, \end{aligned} \tag{4.9}$$

related to the set (4.8) as (4.7) is to (4.6), are linearly independent. They are linear combinations of the equations (4.2) and are equivalent to this latter system in the sense that with every set of values η_{i1}, η_{i2}, ζ_{i1}, ζ_{i2} satisfying equations (4.9) there is associated a unique solution of equations (4.2) whose other elements ξ_1, ξ_2, ϵ_μ are determined by the first, third, and fifth of equations (4.2). The equations (4.4) and (4.9) define a boundary value problem for the $2n$ functions $\eta_i(x)$, $\zeta_i(x)$ analogous to that characterized by the equations (1.1) for the functions $y_i(x)$.

THEOREM 4.1. *For a problem of Bolza having a non-singular normal accessory minimum problem satisfying the non-tangency condition the boundary value problem defined by equations* (4.4) *and* (4.9) *is definitely self-adjoint according to the definition in* §2 *above.*

To prove this theorem we note first that necessary and sufficient conditions for the system (1.1) to be self-adjoint, taken from equations (19) and (20) of the paper I, can be expressed by use of matrix notation in the form

$$TA + \overline{A}T + T' = 0, \quad TB + \overline{B}T = 0, \quad MT^{-1}(a)\overline{M} = NT^{-1}(b)\overline{N}, \tag{4.10}$$

where the bars indicate transposed matrices and T' is the matrix of derivatives of the elements of T. For the boundary value problem defined by equations (4.4) and (4.9) the matrices involved are the $2n$-dimensional matrices

$$A = \begin{pmatrix} V_{ji} & W_{ij} \\ -U_{ij} & -V_{ij} \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & 0 \\ \delta_{ij} & 0 \end{pmatrix},$$

$$M = \begin{pmatrix} \zeta_{i,k1} & -\eta_{i,k1} \\ b_{i,k1} & -a_{i,k1} \end{pmatrix}, \qquad N = \begin{pmatrix} -\zeta_{i,k2} & \eta_{i,k2} \\ -b_{i,k2} & a_{i,k2} \end{pmatrix}.$$

These satisfy the equations (4.10) with the special transformation matrix

$$T = \begin{pmatrix} 0 & \delta_{ik} \\ -\delta_{ik} & 0 \end{pmatrix}.$$

In proving the first equation (4.10) use is made of the symmetry of the matrices U and W, and in proving the third equation relation (4.7) for the various pairs of the solutions (4.8) is needed. The matrix $S = \overline{T}B$ of §2 above is

$$S = \begin{pmatrix} 0 & -\delta_{ij} \\ \delta_{ij} & 0 \end{pmatrix}\begin{pmatrix} 0 & 0 \\ -\delta_{jk} & 0 \end{pmatrix} = \begin{pmatrix} \delta_{ik} & 0 \\ 0 & 0 \end{pmatrix}.$$

Evidently this matrix is symmetric and its quadratic form is non-negative. The only functions $\eta_i(x)$, $\zeta_i(x)$ which make this quadratic form vanish identically have the form $\eta_i(x)\equiv 0$, $\zeta_i(x)$, and no set of functions of this type can satisfy equations (4.4) with $\lambda=0$ and the end conditions (4.9). Otherwise there would be a related solution ξ_1, ξ_2, $\eta_i(x)$, $\mu_\beta(x)$, ϵ_μ of the equations (4.1) and (4.2) with $\eta_i(x)\equiv 0$ on the interval x_1x_2, which is impossible when the accessory minimum problem is normal. Thus all of the conditions (1), (2), (3) of the definition of definite self-adjointness in §2 are satisfied by the boundary value problem associated with equations (4.4) and (4.9), as stated in Theorem 4.1.

The assumption of the non-tangency condition can be omitted, as has recently been suggested to me by W. T. Reid, if the formulation of the accessory minimum problem is slightly modified. The constants ξ_1 and ξ_2 in this problem can be replaced by the values $\eta_{n+1}(x_1)$, $\eta_{n+2}(x_2)$ of two functions $\eta_{n+1}(x)$, $\eta_{n+2}(x)$ subjected to differential equations

$$\eta'_{n+1}=\eta'_{n+2}=0$$

which are to be adjoined to the equations $\Phi_\beta=0$. In the norming integral in the second paragraph of this section the integrand is to be replaced by the sum of the squares of all of the variables $\eta_\sigma(x)$ $(\sigma=1, \cdots, n+2)$. One verifies readily then that for the new problem the end conditions contain only equations of the form of the second, fourth, and last of the equations (4.2) and the construction of the end conditions (4.9) does not involve the non-tangency condition.

5. **Transformations and examples.** If a definitely self-adjoint boundary value problem of the form (1.1) for a set of functions $y_i(x)$ is transformed into one for functions $u_i(x)$ by a non-singular transformation $y_i=U_{ik}(x)u_k$, the property of definite self-adjointness will be preserved. This can be verified by means of the following useful and easily derived transformation formulas, in which the subscript 1 designates the matrices associated with the transformed problem:

$$\begin{aligned} A_1 &= U^{-1}AU - U^{-1}U', & B_1 &= U^{-1}BU, \\ M_1 &= MU(a), & N_1 &= NU(b), \\ P_1 &= U^{-1}(a)P, & Q_1 &= U^{-1}(b)Q, \\ T_1 &= \overline{U}TU, & S_1 &= \overline{U}SU. \end{aligned} \tag{5.1}$$

It is understood that in these formulas a bar indicates a transposed matrix and a prime a matrix of derivatives. With the help of these relations one can readily deduce normal forms for definitely self-adjoint boundary value prob-

lems when the equations involve only two functions y_1 and y_2 and the rank of the matrix $B(x)$, and consequently of $S(x)$, is constant on the interval ab.*

Consider first the case when the determinant of $B(x)$ is everywhere different from zero on the interval ab and the matrix $S(x)$ therefore positive definite. From the second equation (4.10) and the symmetry of $S=\overline{T}B$ it follows that

$$0 = TB + \overline{B}T = (T + \overline{T})B, \qquad 0 = T + \overline{T},$$

so that T is skew-symmetric. There exists a transformation U taking S into the identity matrix, and we have then

$$T = \begin{pmatrix} 0 & t \\ -t & 0 \end{pmatrix}, \qquad S = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

$$B = \begin{pmatrix} 0 & 1/t \\ -1/t & 0 \end{pmatrix}$$

with the help of the relation $S=\overline{T}B$. The transformation

$$U = \begin{pmatrix} 1 & 0 \\ 0 & 1/t \end{pmatrix}$$

now transforms these matrices into the forms shown in (5.3) below, with $s=1/t$. One can readily verify the fact that the most general transformation leaving T and S in (5.3) invariant has the form

$$U = \begin{pmatrix} u_{11} & -s^2 u_{21} \\ u_{21} & u_{11} \end{pmatrix}, \qquad u_{11}^2 + s^2 u_{21}^2 = 1, \tag{5.2}$$

and that it will also leave B in (5.3) invariant under the transformation (5.1). The lower left-hand element of A_1 after such a transformation when set equal to zero, and the derivative of the last equation above, have the forms

$$u_{21}u_{11}' - u_{11}u_{21}' + \cdots = 0,$$

$$u_{11}u_{11}' + s^2 u_{21}\, u_{21}' + ss' u_{21}^2 = 0,$$

where the dots indicate terms not containing derivatives of the elements u_{ik}. If u_{11} and u_{21} are determined by these differential equations with initial values at a single point satisfying the second equation (5.2), they will satisfy that equation identically. The first equation (4.10) shows that $a_{22}=-a_{11}$, and we have the following theorem:

* For a more complete classification see Bamforth, *A classification of boundary value problems for a system of ordinary differential equations of the second order*, Dissertation, University of Chicago, 1927.

THEOREM 5.1. *When $n=2$ every definitely self-adjoint boundary value problem with $|B| \neq 0$ on the interval ab is transformable into a problem with matrices of the form*

$$A = \begin{pmatrix} a_{11} & a_{12} \\ 0 & -a_{11} \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & s^2 \\ -1 & 0 \end{pmatrix},$$
$$T = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \qquad S = \begin{pmatrix} 1 & 0 \\ 0 & s^2 \end{pmatrix} \tag{5.3}$$

with $s(x) \neq 0$ on ab and $|M| = |N|$. Conversely, every problem with these properties is definitely self-adjoint. Such a problem has always an infinity of characteristic numbers.

The relation (20) of I shows that $|M| = |N|$. It is evident that the functions f_i described in Corollary 3.1 above could not all be expansible as there stated if there were only a finite set of characteristic numbers and functions.

The case when the rank of the matrix $B(x)$ is unity everywhere on the interval ab gives rise to a number of normal forms of definitely self-adjoint problems. The matrix $S(x)$ is then transformable into

$$S = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

From the formula $S = \overline{T}B$ and this form of S it can readily be seen that the matrix B and the most general transformation U leaving S invariant have the forms

$$B = \begin{pmatrix} b_{11} & 0 \\ b_{21} & 0 \end{pmatrix}, \qquad U = \begin{pmatrix} \pm 1 & 0 \\ u_{21} & u_{22} \end{pmatrix}.$$

Since B has rank unity the elements b_{11}, b_{21} do not vanish simultaneously, and a transformation U with leading element $+1$ can be chosen so that $u_{22} = -b_{21} + u_{21}b_{11} \neq 0$. Such a transformation will take B into the form

$$B = \begin{pmatrix} b & 0 \\ -1 & 0 \end{pmatrix} \tag{5.4}$$

by means of the second of the formulas (5.1). The conditions $TB + \overline{B}T = 0$, $S = \overline{T}B$, $|T| \neq 0$ then imply that b vanishes identically, and that T has the form

$$T = \begin{pmatrix} t & 1 \\ -1 & 0 \end{pmatrix}. \tag{5.5}$$

With the help of the first equation (19) of I we find that

$$(5.6)\quad A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & -a_{11} \end{pmatrix}, \qquad t = c \exp\left[-\int_a^x 2a_{11}dx\right], \quad a_{12}t = 0.$$

The most general transformation U leaving invariant the matrices S and (5.4) with $b=0$ is found to be

$$U = \begin{pmatrix} \pm 1 & 0 \\ u & \pm 1 \end{pmatrix},$$

and this will also leave T invariant. By means of such a transformation we can make the lower left-hand element in A vanish identically, by a method similar to that used above. The following theorem can then be established without serious difficulty:

THEOREM 5.2. *When $n=2$ every definitely self-adjoint boundary value problem with matrix $B(x)$ identically of rank one on the interval ab is transformable into a problem with matrices of the form*

$$(5.7)\quad A = \begin{pmatrix} a_{11} & a_{12} \\ 0 & -a_{11} \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & 0 \\ -1 & 0 \end{pmatrix}, \qquad S = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

When $a_{12} \not\equiv 0$ the only transformation matrix possible is

$$(5.8)\quad T = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},$$

and the problem is definitely self-adjoint if and only if the end conditions have matrices $M=(m_{ik})$, $N=(n_{ik})$ with $|M| = |N|$.

When $a_{12} \equiv 0$ the possible transformation matrices have the form

$$T = \begin{pmatrix} t & 1 \\ -1 & 0 \end{pmatrix}, \qquad t = c \exp\left[-\int_a^x 2a_{11}dx\right]$$

where c is a constant. The problem is definitely self-adjoint with a matrix T having $c \neq 0$ if and only if the matrices M and N of the end conditions have equal determinants and satisfy the conditions

$$m_{12} = n_{12}\phi, \quad m_{22} = n_{22}\phi, \quad \phi = \exp\left[-\int_a^b a_{11}dx\right], \qquad (m_{12}, m_{22}) \neq (0, 0).$$

The problem is definitely self-adjoint with a matrix T having $c=0$ if and only if M and N have equal determinants and

$$(m_{12} + n_{12}\phi, m_{22} + n_{22}\phi) \neq (0, 0).$$

To prove the second statement of the theorem we note that the last two equations (5.6) imply $t \equiv 0$ when $a_{12} \not\equiv 0$, and that equation (20) of I is satisfied

only when $|M| = |N|$. If these conditions are fulfilled, the problem is self-adjoint. It is also definitely self-adjoint, according to the definition of §3 above, since the equations

$$(5.9) \qquad y_1' = a_{11}y_1 + a_{12}y_2, \qquad y_2' = -a_{11}y_2, \qquad Sy = (y_1, 0) = (0, 0)$$

imply that y_2 must vanish identically when $a_{12} \not\equiv 0$.

For the case when $a_{12} \equiv 0$ the arguments of the preceding paragraphs show that the only possible transformation matrix is (5.5) with t satisfying the conditions in (5.6). The third equation (4.10) implies $|M| = |N|$ and

$$m_{12}^2 - n_{12}^2\phi^2 = m_{12}m_{22} - n_{12}n_{22}\phi^2 = m_{22}^2 - n_{22}^2\phi^2 = 0.$$

For definite self-adjointness the solution

$$y_1 = 0, \qquad y_2 = y_2(a) \exp\left[-\int_a^x a_{11}dx\right]$$

of equations (5.9) must vanish identically if it satisfies also the end conditions of the problem, that is, if

$$(5.10) \qquad (m_{12} + n_{12}\phi)y_2(a) = 0, \qquad (m_{22} + n_{22}\phi)y_2(b) = 0.$$

This will be true if and only if the coefficients of $y_2(a)$ in the last two equations are not both zero. The statements in the theorem now follow readily from equations (5.9) and (5.10)

One can construct without difficulty definitely self-adjoint boundary value problems which have only a finite number of characteristic numbers. For example, the problem with the matrices

$$A = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & 0 \\ -1 & 0 \end{pmatrix}, \qquad M = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad N = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

is definitely self-adjoint with the matrix (5.8) and has the determinant $D(\lambda) = 1$, and hence has no characteristic numbers. The problem with the same matrices A, B and end-matrices

$$M = \begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix}, \qquad N = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}$$

is definitely self-adjoint and has $D(\lambda) = 2 - \lambda(b-a)$. It has a single characteristic number $\lambda = 2/(b-a)$. When

$$M = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad N = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$$

the problem is self-adjoint but not definitely so, and the determinant $D(\lambda)$ vanishes identically. These examples are transforms into the normal forms described above of some equally simple ones communicated to me by Professor W. T. Reid. They show that the property of definite self-adjointness does not imply an infinity of characteristic numbers.

The boundary value problems arising from problems of Bolza in the plane are all of the first type described in Theorem 5.2 and have a_{12} everywhere different from zero. Theorem 3.1 shows that in this case every function $f_1(x)$ with a continuous second derivative on the interval ab is expansible in the form (3.4), provided only that it satisfies the conditions (3.3) with some functions f_2 and g_1 at $x=a$ and $x=b$. It is evident that such problems must have an infinity of characteristic numbers since otherwise such expansions would not be possible in all cases.

UNIVERSITY OF CHICAGO,
CHICAGO, ILL.

SOME EXISTENCE THEOREMS IN THE CALCULUS OF VARIATIONS

I. THE DRESDEN CORNER CONDITION*

BY

E. J. McSHANE

The present note is the first of a series, and is the only one which does violence to the general title by failing to exhibit an existence theorem. Instead, we here perform a few rather easy computations for later use, and by them obtain a simple proof of a corner condition for isoperimetric problems. This corner condition was apparently first established by Dresden as a consequence of the Weierstrass condition, so only the method of proof here can qualify as new. However, it will serve as a suggestive guide to further theorems in later papers.

1. **Notation and continuity assumptions.** In order that some of our preliminary calculations and lemmas shall be valid for both parametric and non-parametric problems, we shall recast non-parametric problems in parametric form. Given an integrand

$$f(x, y, y') \equiv f(x, y^1, \cdots, y^q, y^{1'}, \cdots, y^{q'}),$$

we define

$$\begin{aligned} F(z, z') &= F(z^0, z^1, \cdots, z^q, z^{0'}, z^{1'}, \cdots, z^{q'}) \\ &= z^{0'} f\left(z^0, z^1, \cdots, z^q; \frac{z^{1'}}{z^{0'}}, \cdots, \frac{z^{q'}}{z^{0'}}\right), \qquad z^{0'} > 0, \\ F(z, 0) &= 0. \end{aligned} \tag{1.1}$$

We use a modification of the tensor summation convention. The repetition of a Greek-letter affix in a term requires the summation of the values of that term over all values of the affix. Thus

$$F_{z^\alpha}(z_n, z_n')z^{\alpha'} = F_{z^0}(z_n, z_n')z^{0'} + \cdots + F_{z^q}(z_n, z_n')z^{q'}$$

is summed over $\alpha = 0, 1, \cdots, q$, but not summed over the values of n. Also, if ν is a vector, its length will be denoted by $|\nu|$. Thus if $z = (z^0, \cdots, z^q)$, then $|z| = [z^\alpha z^\alpha]^{1/2}$.

---

* Presented to the Society, December 28, 1937; received by the editors October 29, 1937 and, in revised form, February 16, 1938.

Throughout the papers of this series all integrands $f(x, y, y')$ will be assumed (unless specific statement is made to the contrary) to be defined and continuous together with their partial derivatives of first and second order for all (x, y) in a closed set S, and all y'. Furthermore, we assume without further mention the following:

(1.2) *For every bounded subset S_0 of S there is a constant N such that*

$$f(x, y, y')+N[1+y^{\alpha'}y^{\alpha'}]^{1/2}\geqq 0$$

for all (x, y) in S_0 and all y'.

In the parametric notation, $F(z, z')$ being defined by (1.1), this takes the form:

(1.2′) *For every bounded subset S_0 of S there is a constant N such that $F(z, z')+N|z'|\geqq 0$ for all z in S_0 and all z' with $z^{0'}>0$.*

Parametric integrands $F(z, z')$ (not those arising by (1.1) from non-parametric integrands) will be assumed to be defined and continuous for all z in a closed set S and all z', to be positively homogeneous of degree 1 in z', and to have continuous partial derivatives of the first and second orders for all z in S and for all $z'\neq(0, \cdots, 0)$.

In order to avoid frequent printing of a complicated symbol we make the definitions

$$f_i(x, y, y') \equiv f_{y^{i'}}(x, y, y'), \qquad i = 1, \cdots, q,$$
$$F_i(z, z') \equiv F_{z^{i'}}(z, z'), \qquad i = 0, 1, \cdots, q.$$

From the assumptions made on integrands $f(x, y, y')$ it follows at once that the function $F(z, z')$ defined by (1.1) is continuous with its partial derivatives of first and second orders for all $z\equiv(x, y)$ in S and all z' with $z^{0'}>0$, and is positively homogeneous of degree 1 in z'. Whenever we are discussing the integral of such an integrand along a curve C: $z=z(t)$ it will be tacitly assumed that $z^{0'}(t)>0$ for almost all t, while for all auxiliary curves used in demonstrations we must prove $z^{0'}(t)>0$ for almost all t.

If the representation $z=z(t)$, $(t_1\leqq t\leqq t_2)$, of the curve C satisfies this condition, the function* $F(z(t), \dot{z}(t))$ is measurable. Since by (1.2′) it exceeds a summable function $-N|\dot{z}(t)|$, it has an integral, finite or infinite. Provided that the functions $z^i(t)$ are absolutely continuous, we denote this integral by

$$\mathcal{J}(C) = \int_C F(z, \dot{z})dt = \int_{t_1}^{t_2} F(z(t), \dot{z}(t))dt.$$

* In analogy with Carathéodory's notation, we define $\dot{z}(t)$ to be the vector $(z^{0'}(t), \cdots, z^{q'}(t))$ if the vector is defined and finite; otherwise $\dot{z}(t)$ is defined to be 0 (i.e., $(0, \cdots, 0)$).

Any two absolutely continuous parametrizations of C (subject to the requirement that $z^{0\prime}(t)>0$ for almost all t) give the same value to the integral.* In particular, if $z^0(t)\equiv t$, so that C has the form $y=y(x)$, $(a\leqq x\leqq b)$, and if furthermore the $y^i(x)$ are absolutely continuous, we write

$$\mathcal{J}[y] = \mathcal{J}(C) = \int_a^b f(x, y(x), \dot{y}(x))dx.$$

If $F(z, z')$ is a parametric integrand, defined for all z in a set S and all z', these questions of integrability do not arise. We write in this case also

$$\mathcal{J}(C) = \int_C F(z, \dot{z})dt = \int_{t_1}^{t_2} F(z(t), \dot{z}(t))dt,$$

provided that the functions $z^i(t)$ are absolutely continuous. The invariance under change of parameter is well known in this case. (It also follows from the theorem cited in the preceding footnote.)

2. **Differentiation formulas.** Let $z=z(\sigma)$, $(\sigma_1\leqq\sigma\leqq\sigma_2)$, be a Lipschitzian representation of a curve C. We shall suppose that $z=\phi(\tau)$ and $z=\psi(\tau)$ are two curves passing through the ends of C, so that

$$\phi(\tau_0) = z(\sigma_1), \qquad \psi(\tau_0) = z(\sigma_2), \tag{2.1}$$

and we shall also suppose that ϕ and ψ are absolutely continuous on an interval (a, b) containing τ_0 and have finite derivatives for $\tau=\tau_0$.

From C we form the curve $C(\tau)$ defined by the equation†

$$\begin{aligned} C(\tau)\colon\ z^i = z^i(\sigma, \tau) &= z^i(\sigma) + \frac{\sigma_2 - \sigma}{\sigma_2 - \sigma_1}(\phi^i(\tau) - \phi^i(\tau_0)) \\ &+ \frac{\sigma - \sigma_1}{\sigma_2 - \sigma_1}(\psi^i(\tau) - \psi^i(\tau_0)),\ \sigma_1 \leqq \sigma \leqq \sigma_2,\ i = 0, 1, \cdots, q. \end{aligned} \tag{2.2}$$

This (by (2.1)) joins $\phi(\tau)$ to $\psi(\tau)$, and $z(\sigma, \tau_0)\equiv z(\sigma)$. We wish to calculate the derivative of $\mathcal{J}(C(\tau))$, it being assumed that $C(\tau)$ lies in S for $a\leqq\tau\leqq b$.

Since

$$\mathcal{J}(C(\tau)) = \int_{\sigma_1}^{\sigma_2} F(z(\sigma, \tau), \dot{z}(\sigma, \tau))d\sigma, \tag{2.3}$$

by differentiating under the integral sign we obtain

* E. J. McShane, *Semi-continuity of integrals in the calculus of variations*, Duke Mathematical Journal, vol. 2 (1936), pp. 597–616; in particular, Theorem 2.1 (the first five lines of proof being deleted).

† The curve $C(\tau)$ depends not merely on C and τ, but also on the particular parametrization $z=z(\sigma)$ which we chose for C.

$$\mathcal{J}'(C(\tau))\Big|_{\tau=\tau_0} = \int_{\sigma_1}^{\sigma_2}\left(F_{z^\alpha}\frac{\partial z^\alpha}{\partial\tau} + F_\alpha\frac{\partial\dot z^\alpha}{\partial\tau}\right)d\sigma\Big|_{\tau=\tau_0}$$

$$= \int_{\sigma_1}^{\sigma_2}\left\{F_{z^\alpha}(z(\sigma),\dot z(\sigma))\left[\frac{\sigma_2-\sigma}{\sigma_2-\sigma_1}\phi^{\alpha\prime}(\tau_0) + \frac{\sigma-\sigma_1}{\sigma_2-\sigma_1}\psi^{\alpha\prime}(\tau_0)\right]\right. \tag{2.4}$$

$$\left. + F_\alpha(z(\sigma),\dot z(\sigma))\left[\frac{\psi^{\alpha\prime}(\tau_0)-\phi^{\alpha\prime}(\tau_0)}{\sigma_2-\sigma_1}\right]\right\}d\sigma.$$

If $F(z, z')$ arises from a non-parametric integrand by (1.1), equation (2.4) is still valid if we add the hypothesis that $\dot z^0(\sigma)\geqq m>0$ for almost all σ.

Several special cases of this formula will be of use to us. First, suppose $\phi(\tau)-\phi(\tau_0)=\psi(\tau)-\psi(\tau_0)$, so that (2.2) represents merely a translation of the curve C by the (vector) amount $\phi(\tau)-\phi(\tau_0)$. Then

$$\mathcal{J}'(C(\tau_0)) = \phi^{\alpha\prime}(\tau_0)\int_{\sigma_1}^{\sigma_2}F_{z^\alpha}(z,\dot z)d\sigma. \tag{2.5}$$

(This is independent of the parametric representation of C.)

Suppose next that C is a line segment, and that $C(\tau)$ is the line segment whose ends are $z(\sigma_1)+\tau\pi_1$ and $z(\sigma_2)+\tau\pi_2$, where π_1 and π_2 are given vectors. If the functions $z=z(\sigma)$ representing C are linear, then (2.2) represents the line segment $C(\tau)$ if we take

$$\phi(\tau) = z(\sigma_1)+\tau\pi_1 \quad\text{and}\quad \psi(\tau) = z(\sigma_2)+\tau\pi_2, \tag{2.6}$$

so that (2.4) becomes

$$\mathcal{J}'(C(\tau))\Big|_{\tau=\tau_0} = \int_{\sigma_1}^{\sigma_2}\left\{F_{z^\alpha}(z,\dot z)\left[\frac{\sigma_2-\sigma}{\sigma_2-\sigma_1}\pi_1^\alpha + \frac{\sigma-\sigma_1}{\sigma_2-\sigma_1}\pi_2^\alpha\right]\right. \tag{2.7}$$

$$\left. + F_\alpha(z,\dot z)\left[\frac{\pi_2^\alpha-\pi_1^\alpha}{\sigma_2-\sigma_1}\right]\right\}d\sigma.$$

Applying the mean value theorem to the last term on the right, we obtain

$$\mathcal{J}'(C(\tau))\Big|_{\tau=\tau_0} = \int_{\sigma_1}^{\sigma_2}F_{z^\alpha}(z(\sigma),\dot z(\sigma))\left\{\pi_1^\alpha\frac{\sigma_2-\sigma}{\sigma_2-\sigma_1} + \pi_2^\alpha\frac{\sigma-\sigma_1}{\sigma_2-\sigma_1}\right\}d\sigma \tag{2.8}$$

$$+ F_\alpha(z(\bar\sigma),\dot z(\bar\sigma))\left[\pi_2^\alpha-\pi_1^\alpha\right],$$

where $\sigma_1<\bar\sigma<\sigma_2$.

3. **Interchange of arcs.** Let C_1 and C_2 be rectifiable curves

$$C_1\colon\ z=z_1(\sigma),\ \sigma_1\leqq\sigma\leqq\sigma_2; \qquad C_2\colon\ z=z_2(\tau),\ \tau_1\leqq\tau\leqq\tau_2 \tag{3.1}$$

such that the beginning of C_2 coincides with the end of C_1:

$$z_1(\sigma_2) = z_2(\tau_1). \tag{3.2}$$

The functions $z_1(\sigma)$ and $z_2(\tau)$ will be supposed to satisfy a Lipschitz condition, and if $F(z, z')$ is defined by (1.1) it will also be supposed that except on a set of measure zero $\dot{z}_1^0(t)$ and $\dot{z}_2^0(t)$ are bounded from zero. We define C_{12} to be the curve obtained by traversing first C_1 and then C_2. Furthermore, we define C_{21} to be the curve obtained by starting at $z_1(\sigma_1)$, traversing a curve C_2^* which is a translation of C_2, and then traversing a curve C_1^* which is a translation of C_1. It is clear that the end of C_{21} is the same as the end of C_{12}, namely $z_2(\tau_2)$. We now assume that the point

$$z_1(\sigma) + z_2(\tau) - z_2(t_1) \tag{3.3}$$

lies in S whenever $\sigma_1 \leqq \sigma \leqq \sigma_2$ and $\tau_1 \leqq \tau \leqq \tau_2$, and we proceed to compute $\mathcal{J}(C_{21}) - \mathcal{J}(C_{12})$.

By the definitions,

$$\mathcal{J}(C_{21}) - \mathcal{J}(C_{12}) = \mathcal{J}(C_1^*) - \mathcal{J}(C_1) + \mathcal{J}(C_2^*) - \mathcal{J}(C_2). \tag{3.4}$$

We compute $\mathcal{J}(C_1^*) - \mathcal{J}(C_1)$ by (2.5), where we take $\phi(\tau) = z_2(\tau) - z_1(\sigma_2) + z_1(\sigma_1)$, $\psi(\tau) = z_2(\tau)$. If we momentarily let $C(\tau)$ be the curve (which by hypothesis lies in S), $z = z_1(\sigma) + z_2(\tau) - z_2(\tau_1)$, $(\sigma_1 \leqq \sigma \leqq \sigma_2)$, we obtain by (2.5)

$$\mathcal{J}'(C(\tau)) = \dot{z}_2^a(\tau) \int_{\sigma_1}^{\sigma_2} F_{z^a}(z_1(\sigma) + z_2(\tau) - z_2(\tau_1), \dot{z}_1(\sigma))d\sigma \tag{3.5}$$

for all τ such that $z_2'(\tau)$ exists, that is, for almost all τ. Hence, integrating from $\tau = \tau_1$ to $\tau = \tau_2$, we obtain

$$\begin{aligned} \mathcal{J}(C_1^*) - \mathcal{J}(C_1) &= \mathcal{J}(C(\tau_2)) - \mathcal{J}(C(\tau_1)) \\ &= \int_{\tau_1}^{\tau_2} \int_{\sigma_1}^{\sigma_2} \dot{z}_2^a(\tau) F_{z^a}(z_1(\sigma) + z_2(\tau) - z_2(\tau_1), \dot{z}_1(\sigma))d\sigma d\tau. \end{aligned} \tag{3.6}$$

Next we let $C'(\sigma)$ be the curve $z = z_2(\tau) + z_1(\sigma) - z_1(\sigma_2)$, $(\tau_1 \leqq \tau \leqq \tau_2)$. (Recalling (3.2), we see that this lies by hypothesis in S for $\sigma_1 \leqq \sigma \leqq \sigma_2$.) Then $C'(\sigma_2)$ is C_2; and $C'(\sigma_1)$ is C_2^*, for it is a translation of C_2 and it starts at $z_2(\tau_1) + z_1(\sigma_1) - z_1(\sigma_2)$, which is $z_1(\sigma_1)$ by (3.2). Again using (2.5), but with σ serving in the role of τ, we find

$$\mathcal{J}'(C'(\sigma)) = \dot{z}_1^a(\sigma) \int_{\tau_1}^{\tau_2} F_{z^a}(z_2(\tau) + z_1(\sigma) - z_1(\sigma_2), \dot{z}_2(\tau))d\tau. \tag{3.7}$$

Integrating from σ_1 to σ_2, we have

$$\begin{aligned} \mathcal{J}(C_2) - \mathcal{J}(C_2^*) &= \mathcal{J}(C'(\sigma_2)) - \mathcal{J}(C'(\sigma_1)) \\ &= \int_{\sigma_1}^{\sigma_2} \int_{\tau_1}^{\tau_2} \dot{z}_1^a(\sigma) F_{z^a}(z_2(\tau) + z_1(\sigma) - z_1(\sigma_2), \dot{z}_2(\tau))d\tau d\sigma. \end{aligned} \tag{3.8}$$

With Carathéodory, we now define

$$\Omega(z, p, q) = p^\alpha F_{z^\alpha}(z, q) - q^\alpha F_{z^\alpha}(z, p). \tag{3.9}$$

Then from (3.4), (3.6), and (3.8) we obtain (recalling (3.2))

$$\mathcal{J}(C_{21}) - \mathcal{J}(C_{12}) = -\int_{\tau_1}^{\tau_2}\int_{\sigma_1}^{\sigma_2} \Omega(z_2(\tau) + z_1(\sigma) - z_1(\sigma_2), \dot{z}_1(\sigma), \dot{z}_2(\tau))d\sigma d\tau. \tag{3.10}$$

We digress for a moment to make some remarks on the Ω-function. It is clear that it is continuous for all z, p, and q, if F is the usual parametric integrand. If F arises by (1.1), then Ω is continuous if $p^0>0$ and $q^0>0$. From the definition, $\Omega(z, p, q)$ is positively homogeneous of degree 1 in p and in q:

$$\Omega(z, kp, \kappa q) = k\kappa\Omega(z, p, q) \quad \text{if} \quad k > 0 \quad \text{and} \quad \kappa > 0. \tag{3.11}$$

Also

$$\Omega(z, p, q) = -\Omega(z, q, p). \tag{3.12}$$

Suppose that $F(z, z')$ arises by (1.1) from $f(x, y, y')$, where $(z^0, z^1, \cdots, z^q) = (x, y^1, \cdots, y^q)$, and that the curves C_1, C_2 are represented in the form

$$C_1: \quad y = y_1(x),\ a \leqq x \leqq b; \qquad C_2: \quad y = y_2(x),\ b \leqq x \leqq c,$$

where $y_1(b) = y_2(b)$. We can if we wish revert to the integrand f by simply using x as parameter (it being still assumed that $y_1(x)$ and $y_2(x)$ are Lipschitzian). Then on C_1 we have $\sigma = x$; so $z_1^0(\sigma) = \sigma = x$, and $\dot{z}_1^0(\sigma) \equiv 1$. Likewise $\dot{z}_2^0(\tau) \equiv 1$, so that by (3.10)

$$\begin{aligned}&\mathcal{J}(C_{21}) - \mathcal{J}(C_{12})\\ &= -\int_b^c \left\{\int_a^b \omega(\sigma + \tau - a, y_1(\sigma) + y_2(\tau) - y_1(b), \dot{y}_1(\sigma), \dot{y}_2(\tau))d\sigma\right\} d\tau,\end{aligned} \tag{3.13}$$

where

$$\begin{aligned}\omega(x, y, y', Y') &= f_x(x, y, Y') - f_x(x, y, y')\\ &\quad + y^{\alpha\prime} f_{y^\alpha}(x, y, Y') - Y^{\alpha\prime} f_{y^\alpha}(x, y, y');\end{aligned} \tag{3.14}$$

that is,

$$\omega(x, y, y', Y') = \Omega(x, y; 1, y'; 1, Y'). \tag{3.15}$$

As a special case (useful later) we may assume that C_1 and C_2 are line segments. If $l_i = L(C_i)$ is the length of C_i, $(i = 1, 2)$, we can represent C_1 and C_2 by means of linear functions of arc lengths:

$$C_i: \quad z = z_i(s), \qquad 0 \leqq s \leqq l_i,\ i = 1, 2. \tag{3.16}$$

Then the derivative $\dot{z}_i(s)$ is constantly equal to a unit vector p_i, $(i = 1, 2)$,

and the curves C_{12} and C_{21} bound a parallelogram (which may degenerate into a line segment). Formula (3.10) becomes

$$(3.17)\quad \mathcal{J}(C_{21}) - \mathcal{J}(C_{12}) = -\int_0^{l_1}\left\{\int_0^{l_2}\Omega(z_1(\sigma) + z_2(\tau) - z_2(0), p_1, p_2)d\tau\right\}d\sigma.$$

By the theorem of mean value,

$$(3.18)\quad \mathcal{J}(C_{21}) - \mathcal{J}(C_{12}) = -l_1l_2\Omega(z_1(\bar{\sigma}) + z_2(\bar{\tau}) - z_2(0), p_1, p_2),$$

where $0<\bar{\sigma}<l_1$ and $0<\bar{\tau}<l_2$. If we denote $z_1(\bar{\sigma})+z_2(\bar{\tau})-z_2(0)$ by ξ, then ξ is a point of the parallelogram bounded by C_{12} and C_{21}, and (3.18) becomes

$$(3.19)\quad \mathcal{J}(C_{21}) - \mathcal{J}(C_{12}) = -l_1l_2\Omega(\xi, p_1, p_2).$$

4. **Dresden's corner condition: the parametric form.** In §4 we establish Dresden's corner condition for isoperimetric problems in parametric form.

We shall proceed under the following hypotheses:

(4.1) *The curve* $C: z=z(t)$, $(t_1\leqq t\leqq t_2)$, *is interior to* S *and gives a strong relative minimum to the integral*

$$\mathcal{J}(C) = \int_C F(z, \dot{z})dt$$

in the class of all D' *curves which join* $z(t_1)$ *to* $z(t_2)$ *and give assigned values* γ_j, $(j=1, \cdots, m)$, *to the integrals*

$$\mathcal{G}^j(C) \equiv \int_C G^j(z, \dot{z})dt, \qquad j = 1, \cdots, m.$$

(4.2) *The curve* C *has a corner at* $z(t_0)$.

(4.3) *The curve* C *is normal.** (Hence there is a unique set of multipliers $1, l_1, \cdots, l_m$ such that for the function

$$H(z, z') \equiv F(z, z') + l_\alpha G^\alpha(z, z')$$

the first variation of $\int_C H(z, z)dt$ vanishes along C.)

Then we have the theorem:

THEOREM 1. *Under the hypotheses* (4.1), (4.2), *and* (4.3) *the inequality*†

$$(4.4)\quad \Omega_H(z(t_0), z'(t_0-0), z'(t_0+0)) \leqq 0$$

holds.

* We do not require normality on subarcs of C.

† Here, where several integrands are involved, there are several possible Ω-functions. The subscript indicates the integrand used in defining Ω by (3.9). An analogous notation will be used for the $\mathcal{E}$-function.

For compactness, we define

$$p \equiv z'(t_0 - 0), \qquad r \equiv z'(t_0 + 0). \tag{4.5}$$

Suppose that (4.4) fails to hold; then

$$\Omega_H(z(t_0), p, r) > 0. \tag{4.6}$$

By continuity, there is a positive number δ, so small that C has no corner other than $z(t_0)$ on the arc $z=z(t)$, $t_0-\delta \leqq t \leqq t_0+\delta$, for which

$$\Omega_H(z(\bar{t}), z'(\bar{t}), r) > 0 \quad \text{if} \quad t_0 - \delta \leqq \bar{t} < t_0. \tag{4.7}$$

Define $C_{\tau,0}$ to be the curve obtained from C by interchanging the arcs $z=z(t)$, $(t_0-\delta \leqq t \leqq t_0)$, and $z=z(t)$, $(t_0 \leqq t \leqq t_0+\tau)$, where $0 \leqq \tau \leqq \delta$. The curve $C_{\tau,0}$ is then defined by equations $z=\zeta_\tau(t)$, $(t_1 \leqq t \leqq t_2)$, where

$$\begin{aligned}
&\zeta_\tau(t) = z(t), \quad t_1 \leqq t < t_0 - \delta, \\
&\zeta_\tau(t) = z(t+\delta) - z(t_0) + z(t_0 - \delta), \quad t_0 - \delta \leqq t < t_0 - \delta + \tau, \\
&\zeta_\tau(t) = z(t-\tau) - z(t_0) + z(t_0 + \tau), \quad t_0 - \delta + \tau \leqq t < t_0 + \tau, \\
&\zeta_\tau(t) = z(t), \quad t_0 + \tau \leqq t \leqq t_2.
\end{aligned} \tag{4.8}$$

(Thus $C_{0,0}$ coincides with C. For τ near 0, it is clear that $C_{\tau,0}$ is in S.)

Given any set of m (vector) functions $\eta_j(t) = (\eta_j^\beta(t), \cdots, \eta_j^q(t))$ of class D' on $[t_1, t_2]$ and vanishing at t_1 and t_2, we shall define

$$\begin{aligned}
\Phi(\tau, b) &\equiv \Phi(\tau, b_1, \cdots, b_m) \equiv \mathcal{J}(C_{\tau, b_1, \cdots, b_m}), \\
\Gamma^j(\tau, b) &\equiv \Gamma^j(\tau, b_1, \cdots, b_m) \equiv \mathcal{G}^j(C_{\tau, b_1, \cdots, b_m}),
\end{aligned} \tag{4.9}$$

where $C_{\tau,b}$ is given by the following:

(4.10) *$C_{\tau,b} \equiv C_{\tau, b_1, \cdots b_m}$ is defined by* $z=z(t;\tau,b) \equiv \zeta_\tau(t) + b_\alpha \eta_\alpha(t)$, $(t_1 \leqq t \leqq t_2)$.

For τ and b near 0, this lies in S. As we know,

$$\left.\frac{\partial \Gamma^j}{\partial b_\kappa}\right|_{\tau=b=0} = \int_{t_1}^{t_2} \eta_\kappa^\alpha(t) \left\{ -\frac{d}{dt} G_\alpha^j(z, \dot{z}) + G_{z^\alpha}^j(z, \dot{z}) \right\} dt. \tag{4.11}$$

We have assumed that C is normal. Hence there exist functions $\eta_1(t), \cdots, \eta_m(t)$ of the type described above for which the jacobian

$$\left| \left.\frac{\partial \Gamma^j}{\partial b_\kappa}\right|_{\tau=b=0} \right| \neq 0. \tag{4.12}$$

The equations

$$\Gamma^j(\tau, b) = \gamma_j, \qquad j = 1, \cdots, m, \tag{4.13}$$

have the initial solutions $\tau=b=0$, and at this solution the jacobian with re-

spect to the b_j is not zero by (4.12). So by the implicit functions theorem* there are functions $b=b_\kappa(\tau)$, $(\kappa=1, \cdots, m)$, defined and of class C' for all small non-negative τ, such that

$$\Gamma^j(\tau, b(\tau)) \equiv \gamma_j, \qquad j = 1, \cdots, m. \tag{4.14}$$

Let us define

$$\mathrm{H}(\tau, b) = \Phi(\tau, b) + l_\alpha \Gamma^\alpha(\tau, b) \equiv \int_C H(z, \dot{z})dt. \tag{4.15}$$

Since the first variation of $\int H dt$ vanishes along C, we have

$$\frac{\partial \mathrm{H}}{\partial b_k} = 0 \quad \text{for} \quad \tau = b = 0, \qquad k = 1, \cdots, m. \tag{4.16}$$

By (3.10) and the mean value theorem, for $b=0$ we have

$$\frac{\partial \mathrm{H}}{\partial \tau} = -\int_{t_0-\delta}^{t_0} \Omega_H(z(t), \dot{z}(t), \tau)dt = -\delta\Omega_H(z(\bar{t}), \dot{z}(\bar{t}), \tau) \tag{4.17}$$

for some $\bar{t}$ between $t_0-\delta$ and t_0. Combining (4.16) and (4.17) we have

$$\frac{d}{d\tau}\mathrm{H}(\tau, b(\tau)) = \frac{\partial \mathrm{H}}{\partial \tau} + b_\alpha'(\tau)\frac{\partial \mathrm{H}}{\partial b_\alpha} = \frac{\partial \mathrm{H}}{\partial \tau} < 0 \tag{4.18}$$

for $\tau=0$. But by (4.14) and (4.15),

$$\frac{d}{d\tau}\mathrm{H}(\tau, b(\tau)) = \frac{d}{d\tau}\Phi(\tau, b(\tau)). \tag{4.19}$$

Now by (4.18) and (4.19) we see that

$$\mathcal{J}(C_{\tau,b(\tau)}) = \Phi(\tau, b(\tau)) < \Phi(0, b(0)) = \Phi(0, 0) = \mathcal{J}(C) \tag{4.20}$$

for all small positive τ, while by (4.14)

$$\mathcal{G}^j(C_{\tau,b(\tau)}) = \Gamma^j(\tau, b(\tau)) = \gamma_j. \tag{4.21}$$

This contradicts the minimizing property of C, and establishes the theorem.

5. **Dresden's corner condition: the non-parametric form.** In order that the proof of Theorem 1 shall apply also to integrands $F(z, z')$, $G^j(z, z')$ arising by (1.1), it is only necessary to verify that $z^{0\prime}(t)>0$ along each of the comparison curves $C_{\tau,b}$ used. If C has a class D' representation $z^i=z^i(z^0)$,

* The fact that Γ^j is defined only for $\tau \geqq 0$ does not prevent our use of the theorem; we could for example extend the range of definition of Γ^j as a function of τ and then finally disregard the extension.

$(i=1, \cdots, q)$, then on identifying the parameter t with z^0 we have $\dot{z}^0(t) \equiv 1$ along C. But along $C_{\tau, b(\tau)}$ we have

$$\frac{d}{dt} z^0(t, \tau, b) = \dot{\xi}_\alpha{}^0(t) + b_\alpha \dot{\eta}_\alpha{}^0(t) = 1 + b_\alpha \dot{\eta}_\alpha{}^0(t) \geqq 1 - |b| \max |\dot{\eta}^0(t)| > \tfrac{1}{2}$$

if $|b|$ be small enough. So the comparison curves used, for which $|b|$ is near 0, are allowable, and the proof applies to the non-parametric case also. Recalling (3.15), we have the following result:

THEOREM 2. *Assume the following conditions satisfied:*

(5.1) *The curve $y=y(x)$, $(x_1 \leqq x \leqq x_2)$, is interior to S and gives a strong relative minimum to the integral*

$$\mathcal{J}[y] = \int_{x_1}^{x_2} f(x, y, \dot{y}) dx$$

in the class of all D' curves which join $(x_1, y(x_1))$ to $(x_2, y(x_2))$ and give assigned values γ_j to the integrals

$$\mathcal{G}^j[y] \equiv \int_{x_1}^{x_2} g^j(x, y, \dot{y}) dx, \qquad j = 1, \cdots, m.$$

(5.2) *$y=y(x)$ has a corner at x_0.*

(5.3) *The curve $y=y(x)$ is normal.*

(5.4) *$1, l_1, \cdots, l_m$ are the isoperimetric constants, and*

$$h(x, y, y') \equiv f(x, y, y') + l_\alpha g^\alpha(x, y, y').$$

Then

$$\omega_h(x_0, y(x_0), y'(x_0 - 0), y'(x_0 + 0)) \leqq 0.$$

UNIVERSITY OF VIRGINIA,
CHARLOTTESVILLE, VA.

SOME EXISTENCE THEOREMS IN THE CALCULUS OF VARIATIONS

II. EXISTENCE THEOREMS FOR ISOPERIMETRIC PROBLEMS IN THE PLANE*

BY

E. J. McSHANE

If we seek to find a curve $y=y(x)$, $(x_1 \leqq x \leqq x_2)$, which minimizes an integral $\mathcal{J}[y]=\int f(x, y, \dot{y})dx$ in the class of curves joining two points (x_1, y_1) and (x_2, y_2), a reasonable beginning is to choose a sequence $y=y_n(x)$ for which $\mathcal{J}[y_n]$ tends to the lower bound μ of values of $\mathcal{J}[y]$, and then (under suitable hypotheses) show that a subsequence of the $y_n(x)$ tends uniformly to a limit function $y_0(x)$. However, the uniform convergence of y_n to y_0 does not ensure that $\mathcal{J}[y_n]$ tends to $\mathcal{J}[y_0]$. One way of overcoming this difficulty is to assume that $\mathcal{J}[y]$ is quasi-regular, from which we find $\mathcal{J}[y_0] \leqq \liminf \mathcal{J}[y_n]=\mu$. Since $\mathcal{J}[y_0]$ cannot be less than μ, by the definition of μ, it follows that $\mathcal{J}[y_0]=\mu$. This method of attack was devised by L. Tonelli, and has been applied by him and others, including Graves, Manià, Cinquini, and myself, to a number of different types of variation problems. A second method is to add hypotheses $f(x, y, y')$ which will guarantee that for some sequence $\{y_n(x)\}$ not only does $y_n(x)$ converge uniformly to $y_0(x)$, but also $y_n'(x)$ tends in some manner to $y_0'(x)$, so that $f(x, y_0(x), \dot{y}_0(x))$ is the limit of $f(x, y_n(x), \dot{y}_n(x))$ in some manner which will ensure the convergence of $\mathcal{J}[y_n]$ to $\mathcal{J}[y_0]$. This method does not seem to have been nearly as thoroughly exploited as the first. A very interesting existence theorem, established by this type of reasoning, is to be found in a paper by Hans Lewy.†

Consider now the isoperimetric problem of minimizing $\mathcal{J}[y]$ while keeping $\mathcal{G}[y] \equiv \int g(x, y, y')dx$ constantly equal to a number γ. If we try to use the first of these methods, we find a sequence for which $\mathcal{J}[y] \to \min.$ while $\mathcal{G}[y_n]=\gamma$. But now the limit curve $y=y_0(x)$ must satisfy the equation $\mathcal{G}[y_0]=\gamma=\lim \mathcal{G}[y_n]$, and in order that this shall follow from the uniform convergence of y_n to y_0 the integrand $g(x, y, y')$ must be strongly restricted, in fact, it must be linear in y'. And, in fact, for isoperimetric problems in

* Presented to the Society, April 15, 1938; received by the editors October 29, 1937 and, in revised form, February 16, 1938.

† H. Lewy, *Ueber die Methode der Differenzengleichungen . . .*, Mathematische Annalen, vol. 98 (1928), pp. 107–124.

non-parametric form the only existence theorems* known to me make exactly this requirement on $\mathcal{G}[y]$ or on $\mathcal{J}[y]$.

This suggests turning to the second proof-pattern, and putting conditions on $\mathcal{J}$ and $\mathcal{G}$ which will guarantee that $y_n'(x)$ tends to $y_0'(x)$ in some manner strong enough to ensure that $\mathcal{G}[y_n]$ tends to $\mathcal{G}[y_0]$ and $\mathcal{J}[y_n]$ tends to $\mathcal{J}[y_0]$. In this note (and again in the fourth and fifth of the series) we set forth conditions guaranteeing this.

The method here used is based on a formula (3.18) of I. If Π is a polygon, and two consecutive sides of Π have slopes α, β, respectively, and if it be known that

$$\omega_f(x, y, \alpha, \beta) \geqq 0$$

for all (x, y), then the interchange of these sides does not increase $\mathcal{J}[y]$. Furthermore, if the integrand $g(x, y, y')$ happens to be independent of x and y, the interchange leaves $\mathcal{G}(\Pi)$ unaltered. Suppose then that $\omega_f(x, y, p, r) \geqq 0$ if $p \geqq r$. If we choose a minimizing sequence of polygons $\Pi_n: y = y_n(x)$, we may suppose that for each n the slope of the sides of Π_n increases monotonically. For if ever a side is succeeded by one of lesser slope, we may interchange these sides, leaving $\mathcal{G}$ unaltered and not increasing $\mathcal{J}$. Therefore each function $y_n(x)$ is convex, and when we select a subsequence converging to a limit $y_0(x)$ it will follow that $y_n'(x) \to y_0'(x)$ for almost all x.

In the theorem to be proved, we assume somewhat less than the condition that $\omega_f(x, y, p, r) \geqq 0$ if $p \geqq r$, but the essence of the proof is unchanged; our minimizing sequence is made to consist of several convex arcs.

The notation and definitions used in I will be continued in this note. Also, we add the rather obvious abbreviation "a.c." for "absolutely continuous."

1. **Proofs of some lemmas.** Suppose that $F(z, z')$ is a parametric integrand having the continuity properties required in I, §1, and that C is a rectifiable curve. It is easy to show that if $\{\Pi_n\}$ is a sequence of polygons inscribed in C and having the same initial and final points as C, and the length of the longest side of Π_n tends to 0 as $n \to \infty$, then $\mathcal{J}(\Pi_n) \to \mathcal{J}(C)$. For problems in non-parametric form this is no longer true, as may be shown by examples.†

Suppose, however, that $F(z, z')$ is obtained from a non-parametric integrand $f(x, y, y')$ by (1.1) of I and satisfies the following condition:

---

* L. Tonelli, *Fondamenti di Calcolo delle Variazioni*, vol. 2, pp. 552, 553.

† M. Lavrentieff, *Sur quelques problèmes du calcul des variations*, Annali di Matematica, (4), vol. 4 (1927), pp. 7–28.

L. Tonelli, *Sur une question du calcul des variations*, Matematicheskii Sbornik, vol. 33 (1936), pp. 87–98.

(1.1) *For every bounded subset S_0 of S there are positive numbers δ and a and a number $b \geqq 0$ such that*

$$|F_{zi}(z, z')| \leqq a|z'| + bF(z_0, z')$$

whenever z_0 is in S_0 and $|z - z_0| < \delta$.

Then it can be shown* that if $y = y(x)$, $(a \leqq x \leqq b)$, is a.c., then for every $\epsilon > 0$ there exists a function $y_\epsilon(x)$ having a continuous derivative on $a \leqq x \leqq b$, such that $y_\epsilon(a) = y(a)$ and $y_\epsilon(b) = y(b)$, and such also that

$$\left| \int_a^b f(x, y_\epsilon, \dot{y}_\epsilon)dx - \int_a^b f(x, y, \dot{y})dx \right| < \epsilon.$$

Moreover, if several integrands $f^1, \cdots, f^k$ are such that for each of them (1.1), holds an examination of Tonelli's proof shows that $y_\epsilon(x)$ can be so chosen that

$$\left| \int_a^b f^i(x, y_\epsilon, \dot{y}_\epsilon)dx - \int_a^b f^i(x, y, \dot{y})dx \right| < \epsilon, \qquad i = 1, \cdots, k.$$

In particular, if $f(x, y, y')$ is a function of y' alone, (1.1) surely holds; we need only take $a = \delta = 1$, $b = 0$.

We shall proceed to establish a theorem on isoperimetric problems. Several of the somewhat lengthy hypotheses of this theorem occur again as hypotheses of later theorems, and some of the stages of the proof will also recur; so we separate the hypotheses and split the proof into a sequence of lemmas.

The hypotheses for our first theorem are the following:

(1.2) *The functions $f(x, y, y')$ and $g^j(x, y, y')$, $(j = 1, \cdots, m)$, satisfy the continuity conditions of* I, §1, *on a closed set S_1.*

(1.3) *The functions $f(x, y, y')$ and $g^j(x, y, y')$ satisfy* (1.1), *and on every bounded portion of S the relations*

$$\lim_{|y'| \to \infty} f(x, y, y')/|y'| = \infty,$$

$$\lim_{|y'| \to \infty} g^j(x, y, y')/f(x, y, y') = 0, \qquad j = 1, \cdots, m,$$

hold uniformly in (x, y).

(1.4) *The class K consisting of all a.c. curves $y = y(x)$ joining two fixed points (x_0, y_0) and (X, Y) with $x_0 < X$ and such that the integrals $G^j[y]$ have given fixed values γ_j, $(j = 1, \cdots, m)$, is not empty.*

---

* The proof requires only a minor modification of that given by Tonelli, loc. cit. The condition on F_{xx}, or $f_{x'}$ here stated is superfluous in this connection.

(1.5) *The points* (x_0, y_0) *and* (X, Y) *and the integrals* $\mathcal{J}$, $\mathcal{G}^i$ *have the property that there exist numbers* $a_0, a_1, \cdots, a_m$ *with* $a_0 \geqq 0$ *such that for every number* H *there is a bounded subset* S_H *of* S *containing all a.c. curves* $y = y(x)$ *lying in* S, *joining* (x_0, y_0) *to* (X, Y), *and having*

$$a_0\mathcal{J}[y] + a_\alpha \mathcal{G}^\alpha[y] < H.$$

(1.6) *The symbol* y *stands for a single variable, and the infinite interval* $-\infty < y' < \infty$ *can be subdivided into a finite number of subintervals* $I_1, I_2, \cdots, I_k$ *(not necessarily in that order of precedence from* $-\infty$ *to* ∞*) such that for all* (x, y) *the following relations hold*:

$$\omega_f(x, y, p, r) \geqq 0 \text{ if } p \in I_j \text{ and } r \in I_h, j > h;$$
$$\sigma_j\omega_f(x, y, p, r) \geqq 0 \text{ if } p \in I_j \text{ and } r \in I_j \text{ and } p \geqq r,$$

where $\sigma_j = \pm 1, j = 1, \cdots, k$.

(1.7) *The functions* $g^i(x, y, y')$ *are independent of* x *and* y.

(1.8) S *is the entire* (x, y)*-space.*

(Observe that only in (1.6) do we restrict the number q of variables y^i.) With these hypotheses we can state the following theorem:

THEOREM 1. *Under hypotheses* (1.1), (1.2), (1.3), (1.4), (1.5), (1.6), (1.7), *and* (1.8), *the class* K *contains a curve* $y = y_0(x)$, $(x_0 \leqq x \leqq X)$, *for which* $\mathcal{J}[y]$ *assumes its least value.*

To establish the theorem we first prove six lemmas.

LEMMA 1. *Under hypotheses* (1.2), (1.3), (1.4), *and* (1.5), *the greatest lower bound* μ *of* $\mathcal{J}[y]$ *on the class* K *is finite, and all the curves of* K *for which* $\mathcal{J}[y] < \mu + 1$ *lie interior to a sphere* Q.

First, let C^* be a curve of the class K. By hypothesis (1.5), the subclass K_1 of K on which $\mathcal{J}(C) < \mathcal{J}(C^*) + 1$ lies in a bounded part S_H of S. Let Q be a sphere (including boundary) large enough so that S_H is interior to Q. Then all curves of K_1, and a fortiori all curves C for which $\mathcal{J}(C) < \mu + 1 \leqq \mathcal{J}(C^*) + 1$, lie in Q. This establishes the second statement of the lemma. The g.l.b. of $\mathcal{J}(C)$ on K_1 is clearly the same as its g.l.b. on K. By hypothesis (1.3), there is a number c such that $f(x, y, y') > 1$ if (x, y) is in QS and $|y'| > c$. On the bounded closed set $[(x, y)$ in QS, $|y'| \leqq c]$ the function $f(x, y, y')$ is continuous, hence bounded. So $f(x, y, y')$ is bounded below for all (x, y) in QS and all y'; say $f(x, y, y') \geqq \nu$. It follows that for all curves C of K_1

$$\mathcal{J}(C) \geqq \int_{x_0}^{X} \nu dx = \nu(X - x_0),$$

and $\mathcal{J}(C)$ is bounded below on K_1. Therefore its g.l.b. μ is finite.

LEMMA 2. *If the integrands f, g all satisfy* (1.1) *and hypothesis* (1.4) *holds, then there exists a sequence $\{\Pi_n\}$ of polygons Π_n: $y=y_n(x)$, $(x_0 \leqq x \leqq X)$, such that $y_n(x_0)=y_0$ and $y_n(X)=Y$ for all n, and*

$$\lim_{n\to\infty} \mathcal{J}(\Pi_n) = \mu, \qquad \lim_{n\to\infty} \mathcal{G}^j(\Pi_n) = \gamma_j, \qquad j = 1, \cdots, m.$$

Let $C_1, C_2, \cdots$ be a sequence of curves of K such that $\mathcal{J}(C_n)\to\mu$. Since all the integrands satisfy (1.1), for each n there exists a curve C_n^*: $y=y_n^*(x)$, $(x_0 \leqq x \leqq X)$, joining (x_0, y_0) to (X, Y), having $y_n^{*\prime}$ continuous, such that

$$(1.8) \qquad |\mathcal{J}(C_n^*) - \mathcal{J}(C_n)| < 1/n, \qquad |\mathcal{G}^j(C_n^*) - \mathcal{G}^j(C_n)| < 1/n, \quad j = 1, \cdots, m.$$

Therefore $\mathcal{J}(C_n^*)\to\mu$ and $\mathcal{G}^j(C_n^*)\to\gamma_j$.

If C: $y=y(x)$, $(x_0 \leqq x \leqq X)$, is such that $y'(x)$ is continuous, we form a sequence of inscribed polygons Π_p: $y=y_p(x)$ whose successive vertices are the points

$$(x_0, y_0), (x_0+\delta_p, y(x_0+\delta_p)), \cdots, (x_0+k\delta_p, y(x_0+k\delta_p)), \cdots, (X, Y),$$

where $\delta_p=(X-x_0)/p$. From the continuity of $y'(x)$ it is easily seen that $y_p(x)$ tends uniformly to $y(x)$ and that (neglecting the vertices of the polygons) $y_p'(x)$ tends uniformly to $y'(x)$. Hence for every $\epsilon>0$ there is a Π_p such that $|\mathcal{J}(\Pi_p)-\mathcal{J}(C)|<\epsilon$ and $|\mathcal{G}^j(\Pi_p)-\mathcal{G}^j(C)|<\epsilon$, $(j=1, \cdots, m)$. Applying this to each C_n^*, we see that for each n there is a polygon Π_n: $y=y_n(x)$, $(x_0 \leqq x \leqq X)$, with $y_n(x_0)=y_0$ and $y_n(X)=Y$ for which

$$(1.9) \qquad |\mathcal{J}(\Pi_n) - \mathcal{J}(C_n^*)| < \frac{1}{n}, \quad |\mathcal{G}^j(\Pi_n) - \mathcal{G}^j(C_n^*)| < \frac{1}{n}.$$

From (1.8) and (1.9) we obtain

$$\mathcal{J}(\Pi_n)\to\mu, \quad \mathcal{G}^j(\Pi_n)\to\gamma_j.$$

The lemma is therefore established.

LEMMA 3. *Assume that the following conditions hold:*

(a) *Q is a bounded closed set of points (x, y).*

(b) *The functions $f(x, y, y')$ and $g(x, y, y')$ are continuous functions of their arguments for all (x, y) in Q and all y'.*

(c) *f is non-negative for (x, y) in Q and all y'.*

(d) *For every positive number N there is an $M \geqq 0$ such that if (x, y) is in Q and $|g(x, y, y')| \geqq M$, then $f(x, y, y') \geqq N|g(x, y, y')|$.*

(e) *The a.c. functions $z=z_n(t)$, $(a_n \leqq t \leqq b_n)$, represent a sequence of curves $\{C_n\}$ lying in Q and such that the integrals $\mathcal{J}(C_n)$ are bounded, and the functions*

*$z_n^0(t)$ all satisfy the same Lipschitz condition.**

Then the integrals†

$$\int_E G(z_n(t), \dot{z}_n(t))dt$$

are equi-absolutely continuous, in the sense that for every $\epsilon>0$ there is a $\delta>0$ such that if E is in $[a_n, b_n]$ and $mE<\delta$, then

$$\left|\int_E G(z_n, \dot{z}_n)dt\right| < \epsilon.$$

By hypothesis, there are numbers H, L such that $\mathcal{J}(C_n) \leqq H$ and $\dot{z}_n^0(t) \leqq L$. Let ϵ be a positive number, and let $N=2H/\epsilon$. Hypothesis (d), written in parametric notation, informs us that there is an $M \geqq 0$ such that

$$F(z, \dot{z}) \geqq N\left|G(z, \dot{z})\right| \quad \text{if} \quad z \in Q \quad \text{and} \quad \left|G(z, \dot{z})\right| \geqq M\dot{z}^0.$$

Therefore, since $F \geqq 0$, we have for all $z \in Q$ and all $\dot{z}$ with $\dot{z}^0>0$,

$$F(z, \dot{z}) + MN\dot{z}^0 \geqq N\left|G(z, \dot{z})\right|.$$

In particular, since $L \geqq \dot{z}_n^0(t)>0$ almost everywhere in $[a_n, b_n]$, the inequality

$$F(z_n, \dot{z}_n) + MNL \geqq F(z_n, \dot{z}_n) + MN\dot{z}_n^0 \geqq N\left|G(z_n, \dot{z}_n)\right|$$

holds for almost all t in the interval $[a_n, b_n]$. Now take $\delta=\epsilon/2ML$. If E is in $[a_n, b_n]$ and $mE<\delta$, then

$$\begin{aligned}\left|\int_E G(z_n, \dot{z}_n)dt\right| &\leqq \int_E \left|G(z_n, \dot{z}_n)\right| dt \leqq N^{-1}\int_E [F(z_n, \dot{z}_n) + MNL]dt \\ &\leqq N^{-1}\left[\int_{a_n}^{b_n} F(z_n, \dot{z}_n)dt + MNL \cdot mE\right] \\ &\leqq N^{-1}[H + NML\delta] = \epsilon.\end{aligned}$$

The lemma is therefore established.

Remark. Clearly it would be enough to assume in place of (c) that $f(x, y, y')$ is bounded below for $(x, y) \in Q$ and all y', and to assume in place of (d) that for each $N>0$ there is an $M>0$ and a k independent of M and N such that if $(x, y) \in Q$ and $\left|g(x, y, y')\right| \geqq M$, then $f(x, y, y') \geqq N\left|g(x, y, y')\right| - k$. For if k' is a lower bound for $f(x, y, y')$ for $(x, y) \in Q$, we take K to be the

* This condition on $z_n^0(t)$ certainly holds if $z_n^0(t)=t$; that is, if the C_n are all represented in the form $y=y(x)$, $(a_n \leqq x \leqq b_n)$. As always, we tacitly assume that $z_n^{0\prime}(t)>0$ for almost all t.

† $F(z, \dot{z})$ and $G(z, \dot{z})$ denote, respectively, the parametric integrands associated with $f(x, y, y')$ and $g(x, y, y')$.

larger of k and $-k'$. Then the functions $f(x, y, y')+K$ and $g(x, y, y')$ satisfy the hypotheses of Lemma 3.

LEMMA 4.* *Assume that the following conditions hold:*

(a) *Q is a bounded closed set of points (x, y).*

(b) *$f(x, y, y')$ is continuous for all (x, y) in Q and all y'.*

(c) $\lim_{|y'|\to\infty} f(x, y, y')/|y'| = \infty$ *uniformly for all (x, y) in Q.*

(d) *$\{C_n\}$ is a sequence of a.c. curves $y=y_n(x)$, $(a_n \leqq x \leqq b_n)$, lying in Q and such that the integrals $\mathcal{J}(C_n)$ are bounded.*

Then the functions $y_n(x)$ are equi-absolutely continuous.

The hypotheses of Lemma 3 are satisfied if we take $g(x, y, y') = |y'|$; so the integrals

$$\int_E |\dot{y}_n|\, dx$$

are equi-absolutely continuous. Let ϵ be a positive number. There is a $\delta>0$ such that the integral above is less than ϵ if E is in $[a_n, b_n]$ and $mE<\delta$. If $(\alpha_1, \beta_1), \cdots, (\alpha_k, \beta_k)$ is a set of non-overlapping subintervals of $[a_n, b_n]$ having length $\sum(\beta_i-\alpha_i)<\delta$, then

$$(1.10) \qquad \sum_{i=1}^{k} |y_n(\beta_i) - y_n(\alpha_i)| = \sum_{i=1}^{k} \left| \int_{\alpha_i}^{\beta_i} \dot{y}_n(x)dx \right| \leqq \sum_{i=1}^{k} \int_{\alpha_i}^{\beta_i} |\dot{y}_n(x)|\, dx < \epsilon.$$

This establishes our lemma.

LEMMA 5. *Let hypotheses* (a), (b), (c), *and* (d) *of Lemma 3 be satisfied. Let $\{y_n(x)\}$ be a sequence of a.c. functions all defined on the same interval $[x_0, X]$ and converging everywhere in $[x_0, X]$ to a limit $y_0(x)$. Suppose further that $\dot{y}_n(x)$ tends to $\dot{y}_0(x)$ for almost all x in $[x_0, X]$. Then* $\lim_{n\to\infty} \mathcal{G}[y_n] = \mathcal{G}[y_0]$ *and* $\liminf_{n\to\infty} \mathcal{J}[y_n] \geqq \mathcal{J}[y_0]$. *Moreover, if hypothesis* (c) *of Lemma 4 holds, the limit curve $y=y_0(x)$ is a.c.*

For almost all x we have $y_n(x)\to y_0(x)$ and $\dot{y}_n(x)\to\dot{y}_0(x)$; so for all such x the limit of $f(x, y_n, \dot{y}_n)$ is $f(x, y_0, \dot{y}_0)$, and the limit of $g(x, y_n, \dot{y}_n)$ is $g(x, y_0, \dot{y}_0)$. The function $f(x, y, y')$ is non-negative; therefore by the lemma of Fatou† we obtain

$$\liminf_{n\to\infty} \int_{x_0}^{X} f(x, y_n, \dot{y}_n)dx \geqq \int_{x_0}^{X} f(x, y_0, \dot{y}_0)dx.$$

---

* Lemmas 3 and 4 are closely related to some theorems established by M. Nagumo (*Ueber die gleichmässige Summierbarkeit und ihre Anwendung auf ein Variationsproblem*, Japanese Journal of Mathematics, vol. 6 (1929), pp. 173–182).

† P. Fatou, *Séries trigonométriques et séries de Taylor*, Acta Mathematica, vol. 30 (1906), p. 375

The integrals of the $g(x, y_n, \dot{y}_n)$ are equi-absolutely continuous functions of sets, by Lemma 3; so by a known convergence theorem

$$\lim_{n\to\infty} \int_{x_0}^{X} g(x, y_n, \dot{y}_n)dx = \int_{x_0}^{X} g(x, y_0, \dot{y}_0)dx.$$

The final conclusion is obtained at once from (1.10) if we let $n\to\infty$.

LEMMA 6. *If $y=y_n(x)$, $(x_1\leqq x\leqq x_2)$, is a sequence of real-valued functions defined and convex on the interval $x_1\leqq x\leqq x_2$, and the $y_n(x)$ converge uniformly to a limit function $y_0(x)$ on $x_1\leqq x\leqq x_2$, then $\lim \dot{y}_n(x)=\dot{y}_0(x)$ for almost all x.*

It is easy to see that $y_0(x)$ is also a convex function. Hence each derivative $y_j'(x)$, $(j=0, 1, 2, \cdots)$, is defined almost everywhere on $[x_1, x_2]$ and is monotone increasing. Let E be the set of measure x_2-x_1 on which all these derivatives are defined. Let x_0 be a point of E. If $h>0$ is such that $x+h\leqq x_2$, then

$$\lim_{n\to\infty} \frac{y_n(x_0+h)-y_n(x_0)}{h} = \frac{y_0(x_0+h)-y_0(x_0)}{h}. \tag{1.11}$$

But since $y_n(x)$ is convex, we know that $y_n'(x_0)\leqq [y_n(x_0+h)-y_n(x_0)]/h$. Hence by this and (1.11) we obtain

$$\limsup_{n\to\infty} y_n'(x_0) \leqq \frac{y_0(x_0+h)-y_0(x_0)}{h}, \qquad h>0. \tag{1.12}$$

Now let $h\to 0$; this yields

$$\limsup_{n\to\infty} y_n'(x_0) \leqq y_0'(x_0). \tag{1.13}$$

Repeating the argument with $h<0$, we find

$$\liminf_{n\to\infty} y_n'(x_0) \geqq y_0'(x_0). \tag{1.14}$$

Inequalities (1.13) and (1.14) show that $y_n'(x_0)\to y_0'(x_0)$. Since x_0 is any point of E, this establishes the lemma.

2. **Proof of the theorem; examples.** The preliminaries being disposed of, we take up the proof of the theorem. Let μ be the greatest lower bound of $\mathcal{J}[y]$ on the class K; by Lemma 1 this is finite. By Lemma 2, we can select a sequence of polygons Π_n^*: $y=y_n^*(x)$, $(x_0\leqq x\leqq X)$, joining (x_0, y_0) and (X, Y) and such that

$$\mathcal{J}(\Pi_n^*)\to\mu, \qquad \mathcal{G}^j(\Pi_n^*)=\gamma_{j,n}\to\gamma_j, \qquad j=1, \cdots, m.$$

Suppose that AB and BC are consecutive sides of one of the polygons Π_n^*, having respective slopes α and β. Let D be the fourth vertex of the parallelo-

gram having AB and BC as sides. By "interchanging" AB and BC in the polygon Π_n^* we shall mean (as in I) the operation of forming the polygon Π_n' which has all the sides of Π_n^* except AB and BC, and has sides AD and DC to replace them. From (3.13) or (3.18) of I we know that if $\omega_f(x, y, \alpha, \beta) \geqq 0$ for all (x, y) in the parallelogram $ABCD$, then

$$\mathcal{J}(ADC) - \mathcal{J}(ABC) \leqq 0;$$

whence $\mathcal{J}(\Pi_n') \leqq \mathcal{J}(\Pi_n^*)$. Since by hypothesis (1.7) the integrands $g^j(x, y, y')$ are independent of x and y, it is obvious that $\mathcal{G}^j(\Pi_n') = \mathcal{G}^j(\Pi_n^*)$, $(j = 1, \cdots, n)$.

In particular, if the slope α of AB belongs to an interval I_j and β belongs to an interval I_h with $h < j$, then $\omega_f(x, y, \alpha, \beta) \geqq 0$ for all (x, y), by hypothesis (1.6). So on Π_n^* we search for the first side (that is, side with the least x) whose slope belongs to the interval I_1. If this is not already the first side of Π_n^*, we can interchange it successively with all preceding sides so as to bring it to first place. These interchanges do not increase the integral $\mathcal{J}$ and leave the integrals $\mathcal{G}^j$ unchanged. Next we locate the second one of these sides whose slope belongs to the interval I_1. We can interchange this with preceding sides (if any) so as to bring it to second place; the integral $\mathcal{J}$ is not increased, and the integrals $\mathcal{G}^j$ are unchanged. Proceeding thus, we finally bring all those sides of Π_n^* whose slopes are in I_1 into first, second, $\cdots$, places in unbroken succession. Next we locate the sides whose slopes are in I_2, and bring them together after the sides whose slopes are in I_1; and continue so until all intervals I_h have been considered. We thus have a polygon Π_n' joining (x_0, y_0) to (X, Y), having $\mathcal{J}(\Pi_n')$ not greater than $\mathcal{J}(\Pi_n^*)$, and such that $\mathcal{G}^j(\Pi_n') = \mathcal{G}^j(\Pi_n^*)$, $(j = 1, \cdots, m)$. Moreover, if AB and CD are sides of Π_n' such that the slope of AB is in I_j and the slope of CD is in I_h with $h > j$, then AB precedes CD on the polygon.

Now for any particular number h of the set $1, \cdots, p$ we consider the aggregate of sides of Π_n' whose slopes belong to I_h. Suppose, to be specific, that the number σ_h of (2.5) is $+1$. If AB and BC are consecutive sides of the aggregate having respective slopes α and β, and $\beta < \alpha$, then by hypothesis (1.6) we have $\omega_f(x, y, \alpha, \beta) \geqq 0$ for all (x, y). So, by (3.13) of I, if we interchange AB and BC, the value of $\mathcal{J}$ is not increased. Therefore in this aggregate of sides we seek one having least slope. It can be interchanged successively with all preceding sides so as to bring it to first place in the aggregate; in the interchange the value of $\mathcal{J}$ is not increased and the values of the $\mathcal{G}^j$ are unchanged. Of the remaining sides of the aggregate we seek the one with least slope; this can likewise be brought to second place in the aggregate. Proceeding thus, we find that we can rearrange the sides of the aggregate so as to have their slopes monotonically increasing; the rearrangement leaves the $\mathcal{G}^j$ unchanged

and does not increase the $\mathcal{J}$. If the value of σ_h had been -1, we could have carried out a rearranging process so as to have the slopes of the sides steadily decreasing instead of increasing.

This process having been carried out for each value of h, $(h=1, \cdots, p)$, we arrive finally at a polygon Π_n: $y=y_n(x)$, $(x_0 \leqq x \leqq X)$, joining (x_0, y_0) and (X, Y), having $\mathcal{J}(\Pi_n) \leqq \mathcal{J}(\Pi_n^*)$ and $\mathcal{G}^j(\Pi_n)=\mathcal{G}^j(\Pi_n^*)$, $(j=1, \cdots, m)$. The polygon Π_n consists of at most k polygonal arcs on each of which the slope is either monotonic increasing or monotonic decreasing; that is, Π_n is composed of at most k convex or concave polygonal arcs.

Let us suppose that this has been done for each Π_n^*. For all the polygons Π_n thus obtained the sum

$$a_0\mathcal{J}(\Pi_n) + a_\alpha \mathcal{G}^\alpha(\Pi_n) \leqq a_0\mathcal{J}(\Pi_n^*) + a_\alpha \mathcal{G}^\alpha(\Pi_n^*)$$

is bounded above, since the sequences $\mathcal{J}(\Pi_n^*)$ and $\mathcal{G}^j(\Pi_n^*)$ converge. So by hypothesis (1.5) all the polygons Π_n lie in a bounded set. By Lemma 4, the functions $y_n(x)$ are equi-absolutely continuous, hence equi-continuous; hence by Ascoli's theorem it is possible to select a subsequence which converges uniformly to a limit function $y_0(x)$, $(x_0 \leqq x \leqq X)$. We suppose that this subsequence is the whole sequence $\{y_n(x)\}$.

For each n, we can choose points $x_{0,n}=x_0, x_{1,n}, \cdots, x_{k,n}=X$ such that for $x_{i-1,n} \leqq x \leqq x_{i,n}$ the derivative $y_n'(x)$ is in I_i whenever it is defined. (If y_n' is never in I_i, then $x_{i,n}=x_{i-1,n}$.) It is possible to choose a subsequence of y_n (we suppose it to be the whole sequence) for which $x_{i,n}$ tends to a limit $x_{i,0}$ as $n \to \infty$. On each interval interior to $x_{i-1,0} \leqq x < x_{i,0}$ the convex (or concave) functions $y_n(x)$ tend uniformly to $y_0(x)$; hence $y_0(x)$ is convex (or concave) on the interval $x_{i-1,0} \leqq x < x_{i,0}$. By Lemma 6, on every interval (α, β) interior to $[x_{i-1,0}, x_{i,0}]$ the derivative $\dot{y}_n(x)$ tends almost everywhere to $\dot{y}_0(x)$. Hence the derivative of $y_n(x)$ tends to $\dot{y}_0(x)$ for almost all x in $[x_0, X]$; and by Lemma 5 $y_0(x)$ is a.c. and

$$\mathcal{J}[y_0] \leqq \liminf_{n\to\infty} \mathcal{J}[y_n] \leqq \mu, \quad \mathcal{G}^j[y_0] = \lim_{n\to\infty} \mathcal{G}^j[y_n] = \gamma_j, \quad j = 1, \cdots, m.$$

But $y_0(x)$ is in the class K; so $\mathcal{J}[y_0] \geqq \mu$ by the definition of μ. Hence $\mathcal{J}[y_0]=\mu$, and the theorem is proved.

Examples. (1) An example of a function $f(x, y, y')$ such that $\omega_f(x, y, p, q) > 0$ if $p > q$ is

$$f(x, y, y') = y'^2 + y(1 + y'^2)^{1/2}.$$

Here $f_x=0$, $f_y=(1+y'^2)^{1/2}$; so (1.1) is satisfied with $a=\delta=1$, $b=0$. Also,

$$\omega_f(x, y, p, q) = p(1 + q^2)^{1/2} - q(1 + p^2)^{1/2}$$
$$= [(1 + p^2)(1 + q^2)]^{1/2}\left\{\frac{p}{(1 + p^2)^{1/2}} - \frac{q}{(1 + q^2)^{1/2}}\right\},$$

which is positive if $p > q$. Thus the whole interval $-\infty < y' < \infty$ can be taken as one interval I_1, with $\sigma_1 = +1$, in (1.6). We take, for example, $g^1 = (1+y'^2)^{1/2}$, $g^2 = |y'|^{3/2}$. Hypotheses (1.2) and (1.3) are obviously satisfied. Hypothesis (1.5) holds if we take $a_0 = a_2 = 0$, $a_1 = 1$, for $g^1[y]$ is the length of the curve $y = y(x)$. Let (x_0, y_0) and (X, Y) be any two points with $x_0 < X$, and let γ_1, γ_2 be any two numbers. Then if the class K of a.c. curves joining (x_0, y_0) to (X, Y) and giving $g^1[y]$ and $g^2[y]$ the respective values γ_1, γ_2 is not empty, it contains a minimizing curve for $\mathcal{J}[y]$.

(2) For another example we use the same integral

$$\mathcal{J}[y] = \int_0^1 [\dot{y}^2 + y(1 + \dot{y}^2)^{1/2}]dx,$$

but now we impose no side conditions and require that (x_0, y_0) and (X, Y) be (0, 0) and (0, 1), respectively. Hypotheses (1.2) and (1.3) again hold, as does (1.6). Hypothesis (1.4) is satisfied vacuously. To show that (1.5) also holds (with $a_0 = 1$; there are no other a_i) we use Schwarz' inequality. For $0 \leqq x \leqq 1$ we have

$$(y(x))^2 = \left(\int_0^x \dot{y}(x)dx\right)^2 \leqq x\int_0^x \dot{y}^2dx,$$

$$(y(x))^2 = \left(\int_x^1 \dot{y}(x)dx\right)^2 \leqq (1 - x)\int_0^x \dot{y}^2dx,$$

$$(y(x))^2 \leqq \min(x, 1 - x)\int_0^x \dot{y}^2dx \leqq \frac{1}{2}\int_0^1 \dot{y}^2dx, \qquad \int_0^1 y^2dx \leqq \frac{1}{2}\int_0^1 \dot{y}^2dx.$$

Hence, again using Schwarz' inequality, we have

$$\mathcal{J}[y] = \int_0^1 [(1 + \dot{y}^2) + y(1 + \dot{y}^2)^{1/2}]dx - 1$$
$$\geqq \int_0^1 (1 + \dot{y}^2)dx - \left\{\int_0^1 y^2dx \int_0^1 (1 + \dot{y}^2)dx\right\}^{1/2} - 1$$
$$\geqq (1 - 2^{-1/2})\int_0^1 (1 + \dot{y}^2)dx - 1$$
$$\geqq \frac{1}{4}\int_0^1 [\dot{y}^2 - 4]dx.$$

The last integral satisfies (1.5) (with $a_0=1$); so $\mathcal{J}[y]$ does also. (Note, however, that for pairs of end points other than (0, 0) and (1, 0) the condition (1.5) may fail.)

By Theorem 1 we see that in the class of all a.c. curves joining (0, 0) and (0, 1) there is a minimizing curve for $\mathcal{J}[y]$. This result is not quite trivial; for the integral $\mathcal{J}[y]$ is not even quasi-regular, since

$$f_{y'y'}(x, y, y') = 2 + y(1 + y'^2)^{-3/2},$$

which is not invariant in sign. It is clear that a large class of non-regular problems comes under our theorem, for ω_f involves only the partial derivatives f_x and f_y, and if f satisfies (1.6) so does $f(x, y, y')+\phi(y')$ for every continuously differentiable function $\phi(y')$.

(3) For a third example we take $f(x, y, y')=e^y y'^2$. Here

$$\omega_f(x, y, p, r) = e^y\{r^2 p - p^2 r\} = e^y pr(r - p).$$

So if we take I_1 to be the interval $-\infty < y' < 0$ and I_2 to be $0 \leqq y' < \infty$, we have

$$\omega_f(x, y, p, r) > 0, \qquad p \in I_2, r \in I_1,$$

while

$$-\omega_f(x, y, p, r) > 0, \quad p, r \in I_1 \text{ or } p, r \in I_2, p > r.$$

Hence (1.6) holds with $\sigma_1=\sigma_2=-1$. If, for example, we take $\mathcal{G}^1$ and $\mathcal{G}^2$ as in the first example, the class K contains a minimizing curve for $\mathcal{J}[y]$. This curve will consist either of two concave arcs, on the first of which $y' \leqq 0$ and on the second of which $y' \geqq 0$, or else it will consist of just one concave arc on which y' does not change sign.

3. **Integrals in parametric form.** Related to Theorem 1 there is a theorem on integrals in parametric form; but it applies only to a very restricted class of integrands. The hypotheses on these integrands will be the following:

(3.1) *The function $F(y, x', y')$ is independent of x and of the sign of x':*

$$F(y, -x', y') = F(y, x', y').$$

(3.2) *The functions $G^j(x', y')$, $(j=1, \cdots, m)$, depend only on $|x'|$ and y':*

$$G^j(-x', y') = G^j(x', y').$$

(3.3) *There is a set of constants $a_0, a_1, \cdots, a_m$ with $a_0 \geqq 0$ such that for every number H all the rectifiable curves C beginning at a fixed point and having $a_0\mathcal{J}(C)+a_\alpha\mathcal{G}^\alpha(C) \leqq H$ have lengths less than a constant A_H.*

(3.4) *The interval $-\pi/2 \leqq \theta \leqq \pi/2$ can be subdivided into a finite number of subintervals $I_1, \cdots, I_k$ such that:*

$\Omega_F(x, y, \cos\theta_1, \sin\theta_1, \cos\theta_2, \sin\theta_2) \geqq 0$ *if* $\theta_1 \epsilon I_j$ *and* $\theta_2 \epsilon I_h$ *with* $j > h$,
$\sigma_j\Omega_F(x, y, \cos\theta_1, \sin\theta_1, \cos\theta_2, \sin\theta_2) \geqq 0$ *if* $\theta_1 \epsilon I_j$ *and* $\theta_2 \epsilon I_j$ *and* $\theta_1 \geqq \theta_2$,

where $\sigma_j = \pm 1$, $(j = 1, \cdots, k)$.

(3.5) *The class K of all rectifiable curves C joining two fixed points* (x_0, y_0) *and* (X, Y) *and having* $\mathcal{G}^j(C) = \gamma_j$, *where the* γ_j *are fixed numbers, is a non-empty class.*

THEOREM 2. *Under hypotheses* (3.1), (3.2), (3.3), (3.4), *and* (3.5), *the class K contains a curve* C_0 *for which* $\mathcal{J}(C)$ *is least.*

Let $\{C_n\}$ be a sequence of curves of K for which $\mathcal{J}(C_n)$ tends to the greatest lower bound μ of $\mathcal{J}$ on the class K. Then the numbers $a_0\mathcal{J}(C_n) + a_\alpha\mathcal{G}^\alpha(C_n)$ are bounded; so by (3.3) the lengths $\bar{L}_n$ of the C_n are bounded. Consequently, all the C_n lie in a circle Q. For (x, y) in Q the function $F(y, \cos\theta, \sin\theta)$ is bounded below, say by $-\nu$; hence if C_n is represented with arc length as parameter by functions $x = \bar{x}_n(s)$, $y = \bar{y}_n(s)$, $(0 \leqq s \leqq \bar{L}_n)$, we have

$$\mathcal{J}(C_n) = \int_0^{\bar{L}_n} F(\bar{y}_n, \dot{\bar{x}}_n, \dot{\bar{y}}_n)ds \geqq \int_0^{\bar{L}_n} (-\nu)ds = -\nu\bar{L}_n,$$

and $\mathcal{J}(C_n)$ is bounded below. That is, the number μ is finite.

As we saw in §1, it is possible for each n to construct a sequence of polygons Π_m joining (x_0, y_0) to (X, Y) and tending to C_n such that $\mathcal{J}(\Pi_m) \to \mathcal{J}(C_n)$ and $\mathcal{G}^j(\Pi_m) \to \mathcal{G}^j(C_n)$. We choose for each n one of these polygons (which we rename Π_n) such that

$$|\mathcal{J}(\Pi_n) - \mathcal{J}(C_n)| < 1/n, \qquad |\mathcal{G}^j(\Pi_n) - \mathcal{G}^j(C_n)| < 1/n.$$

Then $\mathcal{J}(\Pi_n) \to \mu$ and $\mathcal{G}^j(\Pi_n) \to \gamma_j$, $(j = 1, \cdots, m)$. We may assume, if we wish, that no side of Π_n is parallel to the y-axis, since we can bring this about by an arbitrarily small change, causing an arbitrarily small change in $\mathcal{J}(\Pi_n)$ and $\mathcal{G}^j(\Pi_n)$.

Suppose that Π_n is defined in terms of arc-length by functions $x = \xi_n(s)$, $y = \eta_n(s)$, $(0 \leqq s \leqq L_n)$. We define a new set of functions x_n, y_n by the relations

$$y_n(s) = \eta_n(s), \qquad x_n(s) = x_0 + \int_0^s |\xi_n(s)|\, ds.$$

Then $y_n(L_n) = Y$ and $x_n(L_n) - x_0 \geqq |X - x_0|$, and

$$\int_0^{L_n} F(y_n, \dot{x}_n, \dot{y}_n)ds = \mathcal{J}(\Pi_n), \qquad \int_0^{L_n} G^j(\dot{x}_n, \dot{y}_n)ds = \mathcal{G}^j(\Pi_n).$$

Exactly as in the preceding proof, we find that the curve $x = x_n(s)$, $y = y_n(s)$

can be considered to be composed of at most k arcs, each one of which is either concave or convex; the jth arc consists of segments along which $(x_n', y_n') = (\cos\theta, \sin\theta)$ with θ in the interval I_j. Along such an arc y_n' is either monotonic increasing or monotonic decreasing. Likewise x_n' is monotonic unless y_n' changes sign; so each of the (at most k) convex or concave arcs can be split into at most two arcs on each of which both x_n' and y_n' are monotonic.

Let us change parameter from s to $t = s/L_n$. The polygon $x = x_n(s)$, $y = y_n(s)$, $(0 \leqq s \leqq L_n)$, is then represented in the form $x = X_n(t)$, $y = Y_n(t)$, $(0 \leqq t \leqq 1)$; and the interval $[0, 1]$ can be split into at most $2k$ subintervals on each of which $X'(t)$ and $Y'(t)$ are monotonic. That is, $X_n(t)$ is concave or convex as a function of t on each separate subinterval, and likewise $Y_n(t)$. The functions $x_n(s)$ and $y_n(s)$ satisfy a Lipschitz condition of constant 1; so $X_n(t)$ and $Y_n(t)$ satisfy a Lipschitz condition of constant L_n, which is bounded. Hence we can select a subsequence (we suppose it the whole sequence) for which $X_n(t)$ and $Y_n(t)$ converge uniformly to limit functions $X_0(t)$ and $Y_0(t)$, respectively. As in the preceding proof, the interval $0 \leqq t \leqq 1$ can be split into at most $2k$ subintervals on each of which $X_0(t)$ is concave or convex, and likewise $Y_0(t)$; and

$$\dot{X}_n(t) \to \dot{X}_0(t), \qquad \dot{Y}_n(t) \to \dot{Y}_0(t)$$

for almost all t.

Since the derivatives of the X_0 and Y_0 are bounded, we have at once

$$\int_0^1 G^j(\dot{X}_0, \dot{Y}_0)dt = \lim_{n\to\infty} \int_0^1 G^j(\dot{X}_n, \dot{Y}_n)dt = \lim_{n\to\infty} \mathcal{G}^j(\Pi_n) = \gamma_j, \qquad j = 1, \cdots, m,$$

$$\int_0^1 F(Y_0, \dot{X}_0, \dot{Y}_0)dt = \lim_{n\to\infty} \int_0^1 F(Y_n, \dot{X}_n, \dot{Y}_n)dt = \lim_{n\to\infty} \mathcal{J}(\Pi_n) = \mu.$$

The curve $x = X_0(t)$, $y = Y_0(t)$ still is not a solution of our problem; for although we have $Y_n(1) = y_n(L_n) = Y$ for all n, so that $Y_0(1) = Y$, still we have only $X_n(1) - x_0 = x_n(L_n) - x_0 \geqq |X - x_0|$, so that $X_0(1) - x_0 \geqq |X - x_0|$. However, let us define t_0 to be that value of t for which

$$x_0 + \int_0^{t_0} \dot{X}_0(t)dt - \int_{t_0}^1 \dot{X}_0(t)dt = X.$$

Such a t_0 exists, for the function

$$x_0 + \int_0^t \dot{X}_0 dt - \int_t^1 \dot{X}_0 dt$$

is a continuous function of t, and as t goes from 0 to 1, it goes from $x_0-|X_0(1)-x_0|$ to $x_0+|X_0(1)-x_0|$, and the number X is not less than the first of these and not greater than the second. If we now define

$$y_0(t) = Y_0(t),$$
$$x_0(t) = x_0 + \int_0^t \sigma(t)\dot{X}_0(t)dt, \qquad 0 \leqq t \leqq 1,$$

where $\sigma(t)=1$ for $0\leqq t\leqq t_0$ and $\sigma(t)=-1$ for $t_0<t\leqq 1$, then $x_0(0)=x_0$, $x_0(1)=X$, $y_0(0)=y_0$, $y_0(1)=Y$; and since $\dot{x}_0(t)=\pm\dot{X}_0(t)$,

$$\int_0^1 F(y_0, \dot{x}_0, \dot{y}_0)dt = \int_0^1 F(Y_0, \dot{X}_0, \dot{Y}_0)dt = \mu,$$
$$\int_0^1 G^j(\dot{x}_0, \dot{y}_0)dt = \int_0^1 G^j(\dot{X}_0, \dot{Y}_0)dt = \gamma_j, \quad j = 1, \cdots, m.$$

Thus the curve $x=x_0(t)$, $y=y_0(t)$ is in the class K and minimizes the integral $\mathcal{J}$.

For an example let us take

$$F = \phi(y)(x'^2 + y'^2)^{1/2}, \quad G^1 = (ax'^2 + by'^2)^{1/2}, \quad G^2 = (x'^4 + y'^4)^{1/4},$$

where a and b are positive and $\phi'(y)>0$ for all y. Conditions (3.1), (3.2), and (3.3) are easily seen to be satisfied. We readily calculate

$$\Omega_F(x, y, \cos\theta_1, \sin\theta_1, \cos\theta_2, \sin\theta_2) = \phi'(y)(\sin\theta_1 - \sin\theta_2),$$

which is positive if $-\pi/2\leqq\theta_2<\theta_1\leqq\pi/2$. Hence (3.4) holds with $k=1$ and $\sigma_1=+1$. By Theorem 2, if (x_0, y_0) and (X, Y) are any two points and γ_1, γ_2 any two numbers such that there are curves C joining (x_0, y_0) to (X, Y) and having $\mathcal{G}^j(C)=\gamma_j$, $(j=1, 2)$, then there is a curve of that class for which $\mathcal{J}(C)$ is least.

UNIVERSITY OF VIRGINIA,
CHARLOTTESVILLE, VA.

CONTRIBUTIONS TO THE THEORY OF MULTIVARIATE STATISTICAL ANALYSIS*

BY

WILLIAM G. MADOW†

The increasing application of statistical methods in the social sciences has resulted in the consideration of hypotheses more complicated that those for which the theory of univariate statistical processes and the elementary theory of multivariate statistical processes suffice. One cause of the complications may be found in the fact that in the social sciences the results of an experiment are frequently vectors of several components rather than of one component. As examples we may note that the prices and quantities of a given commodity in p localities are vectors of p components, and that mental traits are generally tested by batteries of tests and not by a single test. Although the replacement of these vectors by some function, say an average, of their components is adequate for some purposes, yet in certain situations, any such function may be shown to be unsatisfactory. The fact that vector variates must then be analyzed requires the construction of a statistical theory which will facilitate that analysis.

For the normal multivariate distribution, the only distribution for which results have been obtained, the beginnings of that theory may be found in the fairly intensive study, since Gauss, of least squares and multiple correlation, which culminated in the derivation by R. A. Fisher in 1928 of the distribution of the multiple correlation coefficient when the correlation parameter is not zero.‡ The general theory of vector variates has been developed since 1931. It has included the analysis of variance of a vector variate and the theory of relations between sets of variables. Hotelling derived the distribution of the generalization of Student's ratio§ in 1931. Distributions of statistics occurring in the generalized analysis of variance were derived by Wilks‖ in 1932. In 1936 Hotelling stated and solved a general problem in the theory of relations between sets of variables.¶ We continue this analysis.

* Presented to the Society, March 26, 1937; received by the editors October 7, 1937. §2 was read before the Institute of Mathematical Statistics, December 27, 1936; §3 was read before the American Mathematical Society under the title, *The generalized analysis of variance.*

† This research was done under a grant-in-aid from the Carnegie Corporation of New York.

‡ See [13]. The number in the brackets refers to the bibliography at the end of the paper.

§ See [18].

‖ See [34]. Other contributions to the theory are given in [2], [3], [24], [36], [37], [38].

¶ See [20].

In §1 there are given certain definitions and notation which are used throughout the paper.

In §2 we prove several theorems,* the purpose of which is the establishment of a routine method of obtaining the distributions needed in multivariate analysis.

In §3 the theorems of §2 are employed to obtain some of the basic distributions of the theory of vector variates. As illustrations of the uses of the methods, we examine the theory of generalized periodogram analysis,† derive the joint distribution of q^2 and z,‡ and the joint distribution of the two canonical correlations which exist when one of the sets contains two variables and the other set contains at least two variables, under the assumption that not more than one canonical correlation parameter differs from zero. We also analyze the generalized covariance, the tetrad difference,‖ and other statistics occurring in the theory of relation between sets of variables.

It will be noted that the variables are rarely expressed as deviations from a mean or regression function. This is permissible since Theorem 7 enables us to assume that those functions have been eliminated. Thus, if n' is the actual sample number, and if the mean value of the chance variable is a linear function of p parameters, then n, the sample number in terms of which our formulas are expressed, is equal to $n'-p$.

1. **Definitions and notation.** The following definitions and notation will be used throughout this paper unless explicit statement to the contrary is made.

If used as subscripts or superscripts, or as indices of summation or multiplication, the letters i, j, k will take on all integral values from 1 through p; the letters g, h will take on all integral values from 1 through $i-1$, the letter ν will take on all integral values from 1 through n, and the letter γ will take on all integral values from 1 through m. We shall denote by p_γ the sum $n_1+\cdots+n_\gamma$, $(p_0=0)$.

The notation $A\times B$, will be used for the combinatory product¶ of the two sets A and B. We shall write A^2 for $A\times A$. Thus, if we denote by R^1 the totality of all real numbers, then R^ν is the totality of all sets of ν real numbers. If V is a subset of $A\times B$ and b is an element of B, we shall denote by $V(b)$

* These are more fully described in the first few paragraphs of §2.

† For a discussion of the Schuster periodogram in one variable, see [14] and [17].

‡ See [20], p. 333. These statistics are measures of the degree of dependence of sets of chance variables.

‖ M. A. Girshick has obtained some of these distributions under slightly less general conditions by other methods. They have not, as yet, been published.

¶ That is, $A\times B$ is the set consisting of all pairs (a, b) where a is an element of A and b is an element of B.

the totality of elements a of A, such that (a, b) is an element of V.

If $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$ are chance variables* with the distribution function† $P(\xi_1, \xi_2, \cdots, \xi_n)$, and if

$$(1.1)\quad P(\xi_1, \xi_2, \cdots, \xi_n) = \int_{-\infty}^{\xi_n} \cdots \int_{-\infty}^{\xi_2} \int_{-\infty}^{\xi_1} f(x_1, x_2, \cdots, x_n) dx_1 dx_2 \cdots dx_n,$$

where $f(x_1, \cdots, x_n)$ is a real single-valued non-negative summable function of the n real variables $x_1, \cdots, x_n$ defined on R^n, then $f(x_1, \cdots, x_n)$ is called the probability density of the chance variables $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$. If $n = p_m$ and

$$P(\xi_1, \cdots, \xi_n) \equiv \prod_{\gamma} P(\xi_{p_{\gamma-1}+1}, \cdots, \xi_{p_\gamma}),$$

then the sets of chance variables $\mathbf{x}_{p_{\gamma-1}+1}, \cdots, \mathbf{x}_{p_\gamma}$ will be said to be independently distributed. The symbol $\mathcal{D}(\mathbf{x}_1, \cdots, \mathbf{x}_n)$ will stand for "the probability density of the chance variables‡ $\mathbf{x}_1, \cdots, \mathbf{x}_n$."

It is easy to see that if a set of chance variables has a probability density, then any subset of them will have a probability density. Furthermore if $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$ have a probability density and if $n = p_m$, then a necessary and sufficient condition that the sets of chance variables $\mathbf{x}_{p_{\gamma-1}+1}, \cdots, \mathbf{x}_{p_\gamma}$ be independently distributed is

$$\mathcal{D}(\mathbf{x}_1, \cdots, \mathbf{x}_n) \equiv \prod_{\gamma} \mathcal{D}(\mathbf{x}_{p_{\gamma-1}+1}, \cdots, \mathbf{x}_{p_\gamma}).$$

If $\phi(x_1, \cdots, x_n)$ is a measurable function§ of $x_1, \cdots, x_n$ on R^n, then the symbol $\mathcal{E}\{\phi(\mathbf{x}_1, \cdots, \mathbf{x}_n)\}$ will stand for "the mathematical expectation of $\phi(\mathbf{x}_1, \cdots, \mathbf{x}_n)$," and, by definition,

$$\mathcal{E}\{\phi(\mathbf{x}_1, \cdots, \mathbf{x}_n)\} = \int \cdots \int_{R^n} \phi(x_1, \cdots, x_n) dP(x_1, \cdots, x_n),$$

---

* For definitions of chance variables, see [7], p. 8 and [21], p. 20. Unfortunately there are differences in the terminology used in papers in statistics and in probability. To avoid confusion, one must remember that variate, statistic, and chance variable are synonyms, as are probability density and distribution. Writers on statistical subjects sometimes use the term cumulated distribution for distribution function. It is particularly to be noted that the term distribution function in the literature of the theory of probability is the integral of the distribution referred to in papers on the theory of statistics.

† For all sets of real numbers $\xi_1, \xi_2, \cdots, \xi_n$, the probability that $\mathbf{x}_1 < \xi_1, \mathbf{x}_2 < \xi_2, \cdots, \mathbf{x}_n < \xi_n$ is given by $P(\xi_1, \xi_2, \cdots, \xi_n)$. For a discussion of probability density and distribution function, see [7] and [21].

‡ In using $\mathcal{D}$ we, of course, assert that $\mathbf{x}_1, \cdots, \mathbf{x}_n$ are chance variables and have a probability density.

§ We will usually extend the domain of definition of functions defined on a subset of R^r to cover the set R^r by setting the functions equal to zero outside their original domain of definition.

where the integral on the right-hand side is a Lebesgue-Stieltjes integral.*

The positive definite quadratic forms q and q_ν will be defined by the equations

$$q = \sum_{i,j} \sigma^{ij} x_i x_j$$

and

$$q_\nu = \sum_{i,j} \sigma^{ij} x_{i\nu} x_{j\nu},$$

respectively. The matrix of q and q_ν will be denoted by (σ^{-1}). The determinant of (σ^{-1}) will be denoted by σ^{-1}. The inverse of (σ^{-1}) will be written (σ), and the elements of (σ) will be denoted by σ_{ij}, where σ_{ij} is the cofactor of σ^{ij} in σ^{-1} divided by σ^{-1}. The determinant of (σ) is, of course, σ.

The function $N(x; c)$ will be defined by the equations

$$N(x; c) = (2\pi c)^{-1/2} \exp\left[-x^2/2c\right], \qquad -\infty < x < \infty; 0 < c,$$
$$N(x; c) = 0, \qquad -\infty < x < \infty; c \leqq 0.$$

The functions $N((x); (\sigma))$ and $N((x_\nu); (\sigma))$ will be defined by the equations

$$N((x); (\sigma)) = (2\pi\sigma)^{-1/2} \exp\left[-q/2\right], \qquad -\infty < x_i < \infty,$$
$$N((x_\nu); (\sigma)) = (2\pi\sigma)^{-1/2} \exp\left[-q_\nu/2\right], \qquad -\infty < x_i < \infty.$$

Of course the elements σ_{ij} may assume any values as long as the requirement that q and q_ν are positive definite is satisfied.

If $\mathcal{D}(\mathbf{x}) = N(x; \sigma^2)$, then $\mathbf{x}$ will be said to be normally distributed with variance σ^2 or to have a normal distribution with variance σ^2. If $\mathcal{D}(\mathbf{x}_1, \cdots, \mathbf{x}_p) = N((x); (\sigma))$, then $\mathbf{x}_1, \cdots, \mathbf{x}_p$ will be said to have a normal multivariate distribution.

The function $G(x; n, c)$ will be defined by the equations

$$G(0; 0, c) = 1, \qquad -\infty < c < \infty,$$
$$G(x; n, c) = \left[\Gamma(n/2)\right]^{-1}(2c)^{-n/2}x^{n/2-1} \exp\left[-x/2c\right],$$
$$0 < x < \infty; 0 < n < \infty; 0 < c < \infty,$$

and $G(x; n, c) = 0$ for all other values of x, n, c. The function $\overline{G}(\lambda; n, c)$ will be defined by the equation,

$$\overline{G}(\lambda; n, c) = \int_0^\infty x^{\lambda/2} G(x; n, c)dx, \qquad \lambda \geqq -n + 1.$$

* If a measurable function $y = \phi(x_1, \cdots, x_n)$ has been defined, we shall mean by $\mathbf{y}$ the function $\phi(\mathbf{x}_1, \cdots, \mathbf{x}_n)$.

Then it follows that

$$\overline{G}(\lambda; n, c) = (2c)^{\lambda/2} \frac{\Gamma((n+\lambda)/2)}{\Gamma(n/2)}.$$

If $\mathcal{D}(x) = G(x; n, 1)$, then x will be said to have a χ^2 distribution with n degrees of freedom. In that case the distribution of $x/2$ will be said to be an incomplete Γ-function distribution of index n.

The function $D((x); (n))$ will be defined by the equations

$$D((x); (n)) = \frac{\Gamma(n/2)}{\prod_{i=1}^{p+1}\Gamma(n_i/2)}\left(\prod_{i=1}^{p} x_i^{n_i/2-1}\right)(1 - x_1 - \cdots - x_p)^{n_{p+1}/2-1},$$
$$0 < x_i;\ x_1 + \cdots + x_p < 1,\ 1 \leqq n_i,\ n = n_1 + \cdots + n_{p+1},$$

and $D = 0$ otherwise. The function $\overline{D}((\lambda); (n))$ is defined by the equation,

$$\overline{D}((\lambda); (n)) = \int_{0<x_p} \cdots \int_{0<x_1}^{x_1+\cdots+x_p<1} \left(\prod_{i=1}^{p} x_i^{\lambda_i/2}\right) D((x); (n)) dx_1 \cdots dx_p.$$

Then letting $\lambda = \lambda_1 + \cdots + \lambda_p$, we find that

$$\overline{D}((\lambda); (n)) = \frac{\Gamma(n/2)}{\Gamma((n+\lambda)/2)} \cdot \prod_{i=1}^{p} \frac{\Gamma((n_i+\lambda_i)/2)}{\Gamma(n_i/2)}.$$

If $\mathcal{D}(x_1, \cdots, x_p) = D((x); (n))$, then $x_1, \cdots, x_p$ will be said to have an incomplete Dirichlet distribution of indices $n_1, \cdots, n_{p+1}$. If $p = 1$, then x_1 and x_2 will be said to have an incomplete β-function distribution of indices n_1 and n_2.

2. **General theorems.** One of the most important theorems of univariate statistical analysis is that which states necessary and sufficient conditions that a set of quadratic forms in normally and independently distributed chance variables be themselves independently distributed with a joint probability density which is a product of χ^2 distributions.*

The fundamental part of this theorem may be stated as follows:

THEOREM 1. (Cochran's theorem.) *If $x_1, x_2, \cdots, x_n$ are normally and independently distributed, with zero means and unit variances, if $q_1, q_2, \cdots, q_m$ are quadratic forms in $x_1, x_2, \cdots, x_n$, with constant coefficients such that*

$$\sum_{\gamma} q_\gamma = \sum_{\nu} x_\nu^2, \tag{2.1}$$

and if the rank of q_γ is n_γ, then a necessary and sufficient condition that there exist n linear functions with constant coefficients,

* The theorem is stated by W. G. Cochran [6], p. 178, and essentially stated by R. A. Fisher [12], p. 97.

$$z_\mu = \sum_\nu c_{\mu\nu} x_\nu, \qquad \mu = 1, \cdots, n, \tag{2.2}$$

which are normally and independently distributed with zero means and unit variances, and are such that

$$\sum_{\nu=p_{\gamma-1}+1}^{p_\gamma} z_\nu^2 = q_\gamma, \tag{2.3}$$

*is**

$$p_m = n. \tag{2.4}$$

The necessity of the condition (2.4) is obvious. For a detailed proof of the sufficiency of the condition, the reader is referred to [6], p. 178. It may be remarked, however, that the proof depends on the fact that in reducing the q_γ's to sums of squares, the n linear functions thus obtained must be linearly independent or contradict (2.1).

In this section we prove several theorems by means of which many problems in multivariate statistical analysis may be solved. Of these, Theorems 7, 8, and 9 are generalizations of the Fisher-Cochran theorem.

A problem of statistical analysis may be formulated in the following way: If the distribution of several functions of certain chance variables has been derived, then what is the distribution of the same functions of other chance variables?

As an example of this problem we may cite the extension of the distribution of the multiple correlation coefficient of variables having a normal multivariate distribution with zero multiple correlation parameter† to the distribution of the multiple correlation coefficient of variables having a normal multivariate distribution, the multiple correlation parameter of which need not vanish.‡

In Theorem 2 the problem is solved under fairly general conditions. In particular the problem is solved when the functions are sufficient statistics§ for both sets of chance variables. Various properties of sufficient statistics are then studied in Theorems 3 and 4 in order to determine to what extent

---

* The algebraic content of this theorem may be stated as follows: *If the real quadratic forms* $q_1, q_2, \cdots, q_m$ *in* $x_1, x_2, \cdots, x_n$ *are such that* $\sum_\gamma q_\gamma = \sum_\nu x_\nu^2$, *and if the rank of* q_γ *is* n_γ, *then a necessary and sufficient condition that there exist an orthogonal transformation* $z_\mu = \sum_\nu c_{\mu\nu} x_\nu$, $(\mu = 1, \cdots, n)$, *such that* $\sum_{\nu=p_{\gamma-1}+1}^{p_\gamma} z_\nu^2 = q_\gamma$, *is* $p_m = n$.

† This distribution was derived by R. A. Fisher. See [10], p. 91 and [11], p. 811.

‡ This distribution was derived by Fisher four years after the appearance of the preceding distribution. See [13], p. 660.

§ See [28], §4, for a general definition of a system of sufficient statistics.

the possession of a set of sufficient statistics defines the underlying chance variables and to derive an expression for the probability density of any system of sufficient statistics.

In Theorem 5 we show that certain useful transformations may be treated as though they were the results of iterating linear transformations.

Frequently, in statistical investigations one is confronted with the necessity of finding distributions subject to the condition that certain variables are held constant or the condition that the effect of certain variables be previously eliminated. In such cases one must learn if the distributions obtained are functions of the values of the variables held constant and of the distribution of these variables. In Theorem 6 we shall provide a means of obtaining this information in situations which essentially depend on the normal distribution.

The following well known lemma will be used in the proofs of Theorems 2 and 4:

LEMMA 1. *Let the real single-valued Borel-measurable* functions*

$$y_i = g_i(x_1, \cdots, x_n) \tag{2.5}$$

be defined on R^n, and let $\psi(y_1, \cdots, y_p)$ be a single-valued measurable function of $y_1, \cdots, y_p$ defined on R^p. Upon substituting from (2.5), it is seen that $\psi(y_1, \cdots, y_p)$ is a single-valued measurable function, $\phi(x_1, \cdots, x_n)$, of $x_1, \cdots, x_n$.

Then, if $\mathbf{x}_1, \cdots, \mathbf{x}_n$ are chance variables, it follows that $\mathbf{y}_1, \cdots, \mathbf{y}_p$, where $\mathbf{y}_i = g_i(\mathbf{x}_1, \cdots, \mathbf{x}_n)$, are chance variables and

$$\mathcal{E}\{\psi(\mathbf{y}_1, \cdots, \mathbf{y}_p)\} = \mathcal{E}\{\phi(\mathbf{x}_1, \cdots, \mathbf{x}_n)\}.$$

A proof of this Lemma is given in [21], p. 35. A similar result is true when the functions (2.5) are measurable. It is noted that if

$$\psi(y_1, \cdots, y_p) = \exp\left[i \sum_j t_j \mu_j(y_1, \cdots, y_p)\right],$$

where μ_j is a real single-valued measurable function of $y_1, \cdots, y_p$, then we have the following equation, which is often used in applying the Fourier transform theory in the derivation of probability densities,

$$\mathcal{E}\left\{\exp\left[i \sum_j t_j \lambda_j(\mathbf{x}_1, \cdots, \mathbf{x}_n)\right]\right\} = \mathcal{E}\left\{\exp\left[i \sum_j t_j \mu_j(\mathbf{y}_1, \cdots, \mathbf{y}_p)\right]\right\},$$

* For a definition and discussion of Borel-measurable functions, see, for example, [31], chaps. 1, 2, 3.

where $\lambda_j(x_1, \cdots, x_n)$ is the real single-valued measurable function of $x_1, \cdots, x_n$ obtained by substituting from (2.5) in μ_j.

THEOREM 2. *Let*

$$\mathfrak{D}(\mathbf{x}_1, \cdots, \mathbf{x}_n) = f(x_1, \cdots, x_n),$$

and let

$$\mathfrak{D}(\mathbf{x}_1', \cdots, \mathbf{x}_n') = F(x_1', \cdots, x_n'),$$

where F vanishes whenever f vanishes. Let the real single-valued Borel-measurable functions $y_1, \cdots, y_p$ *be defined as in* (2.5), *and let* $y_1', \cdots, y_p'$ *be the same functions of* $x_1', \cdots, x_n'$ *that* $y_1, \cdots, y_p$ *are of* $x_1, \cdots, x_n$. *Suppose that, when we substitute for* y_i *and* y_i',

$$f(x_1, \cdots, x_n) = k(y_1, \cdots, y_p) \cdot l(x_1, \cdots, x_n) \tag{2.6}$$

and

$$F(x_1', \cdots, x_n') = K(y_1', \cdots, y_p') \cdot l(x_1', \cdots, x_n'). \tag{2.7}$$

Then if

$$\mathfrak{D}(\mathbf{y}_1, \cdots, \mathbf{y}_p) = h(y_1, \cdots, y_p),$$

it follows that

$$\mathfrak{D}(\mathbf{y}_1', \cdots, \mathbf{y}_p') = \frac{K(y_1', \cdots, y_p') \cdot h(y_1', \cdots, y_p')}{k(y_1', \cdots, y_p')},$$

where $\mathfrak{D}(\mathbf{y}_1', \cdots, \mathbf{y}_p') = 0$, *whenever* $k = 0$.

Proof. For any Borel-measurable set A of R^p, let

$$\psi_A(y_1', \cdots, y_p') = \frac{K(y_1', \cdots, y_p')}{k(y_1', \cdots, y_p')}$$

if $y_1', \cdots, y_p'$ is an element of the set A and $k(y_1', \cdots, y_p') \neq 0$, and let $\psi_A(y_1', \cdots, y_p') = 0$ otherwise. Also let

$$\phi_{A^{-1}}(x_1', \cdots, x_n') = \frac{F(x_1', \cdots, x_n')}{f(x_1', \cdots, x_n')}$$

if $x_1', \cdots, x_n'$ is an element of A^{-1} and $f(x_1', \cdots, x_n') \neq 0$, where A^{-1} is the set of values of $x_1', \cdots, x_n'$ such that* $y_1', \cdots, y_p'$ is an element of A, and let $\phi_{A^{-1}}(x_1', \cdots, x_n') = 0$ otherwise.

Then, from Lemma 1, for all Borel-measurable sets A of R^p

* The fact that $y_1', \cdots, y_p'$ are Borel-measurable functions of $x_1', \cdots, x_n'$ causes A^{-1} to be a Borel set of R_n.

$$\int \cdots \int_A \frac{K(t_1, \cdots, t_p) \cdot h(t_1, \cdots, t_p)}{k(t_1, \cdots, t_p)} dt_1 \cdots dt_p = \int \cdots \int_{A^{-1}} F(x_1, \cdots, x_n) dx_1 \cdots dx_n.$$

Since the integral on the right is equal to the probability that $y_1', \cdots, y_p'$ is an element of A, the theorem follows from (1.1).

As an example of the possible uses of Theorem 2, let us consider the derivation of the joint distribution of variances and covariances of chance variables which have a normal multivariate distribution.*

Hotelling† has, by a simple geometric argument, obtained the joint distribution of sample correlation coefficients when all the "true" correlation coefficients are zero. In this case, it is known that the variances are independent of the correlation coefficients.‡ We may therefore state Hotelling's result more generally as follows:

Let $\mathcal{D}(x_{11}, \cdots, x_{pn}) = \prod_i \prod_\nu N(x_{i\nu}; 1)$. If

$$a_i = \sum_\nu x_{i\nu}^2; \qquad r_{ij} = \frac{\sum_\nu x_{i\nu} x_{j\nu}}{a_i^{1/2} a_j^{1/2}},$$

then

$$\mathcal{D}(\boldsymbol{a}_1, \cdots, \boldsymbol{a}_p, \boldsymbol{r}_{12}, \cdots, \boldsymbol{r}_{p-1,p}) = \pi^{-p(p-1)/4} \omega^{(n-p+1)/2-1} \prod_j G(a_j; n - j + 1, 1),$$

where

$$\omega = \begin{vmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{12} & 1 & \cdots & r_{2p} \\ \cdot & \cdot & \cdots & \cdot \\ r_{1p} & r_{2p} & \cdots & 1 \end{vmatrix}.$$

We are able, by means of Theorem 2, to extend Hotelling's result to dependent chance variables.

Let $\mathcal{D}(x_{11}', \cdots, x_{pn}') = \prod_\nu N((x_\nu); (\sigma))$. If a_i', r_{ij}', and ω' are the same functions of $x_{11}', \cdots, x_{pn}'$ that a_i, r_{ij}, and ω are of $x_{11}, \cdots, x_{pn}$, then

$$\mathcal{D}(\boldsymbol{a}_1', \cdots, \boldsymbol{a}_p', \boldsymbol{r}_{12}', \cdots, \boldsymbol{r}_{p-1,p}') = 2^{-pn/2} \pi^{-p(p-1)/4} \sigma^{-n/2} \omega'^{(n-p+1)/2-1}$$

$$\cdot \left\{ \prod_j [\Gamma((n - j + 1)/2)]^{-1} a_j'^{n/2-1} \right\} \cdot \exp\left[-\frac{1}{2} \sum_{i,j} \sigma^{ij} r_{ij}' a_i'^{1/2} a_j'^{1/2} \right]. \tag{2.8}$$

* This distribution was first found by J. Wishart [39], p. 38, and [4], p. 270.

† See [19], p. 519.

‡ An analytic proof of this statement and Hotelling's distribution is readily obtained by means of Theorem 8.

This conclusion is immediate when one considers that in the notation of Theorem 2, $l(x_{11}, \cdots, x_{pn}) = l(x'_{11}, \cdots, x'_{pn}) \equiv 1$,

$$k(a_1, \cdots, a_p, r_{12}, \cdots, r_{p-1,p}) = (2\pi)^{-pn/2} \exp\left[-\tfrac{1}{2}(a_1 + \cdots + a_p)\right]$$

and

$$K(a_1, \cdots, a_p, r_{12}, \cdots, r_{p-1,p}) = (2\pi)^{-pn/2}\sigma^{-n/2} \exp\left[-\frac{1}{2}\sum_{i,j}\sigma^{ij}r_{ij}a_i^{1/2}a_j^{1/2}\right].$$

The transformation into variances and covariances is obvious.

The chief reason for the introduction of the following theorems on sufficient statistics is the desire to learn to what extent the requirements that (2.6) and (2.7) are satisfied restrict the applicability of Theorem 2. We will show, in the corollary to Theorem 3, that at least for sufficient statistics, using the definition based on (2.9), the chief value of Theorem 2 is that it permits the extension of probability densities derived on the assumption that certain parameters have fixed values, to probability densities which are functions of the values of those parameters just as in the preceding example.

Let the chance variable $\mathbf{x}_\nu$ have a probability density $f(x_\nu, \theta_1, \cdots, \theta_q)$, $(q<n)$, which depends on several parameters, $\theta_1, \cdots, \theta_q$, and let the chance variables $\mathbf{x}_1, \cdots, \mathbf{x}_n$ be independently distributed. Let the real single-valued Borel-measurable functions $y_1, \cdots, y_p$ be defined as in (2.5). If

$$(2.9) \qquad \prod_\nu f(x_\nu, \theta_1, \cdots, \theta_q) = k(y_1, \cdots, y_p, \theta_1, \cdots, \theta_q) \cdot l(x_1, \cdots, x_n),$$

where $k(y_1, \cdots, y_p, \theta_1, \cdots, \theta_q)$ and $l(x_1, \cdots, x_n)$ are real single-valued measurable functions of their indicated variables, then $\boldsymbol{y}_1, \cdots, \boldsymbol{y}_p$ are called a sufficient system of statistics* with respect to the estimation of $\theta_1, \cdots, \theta_q$.

Now let the chance variable $\mathbf{x}'_\nu$ have a probability density $F(x'_\nu, \lambda_1, \cdots, \lambda_s)$ which depends on several parameters $\lambda_1, \cdots, \lambda_s$, and let the chance variables $\mathbf{x}'_1, \cdots, \mathbf{x}'_n$ be independently distributed.

The question suggested by (2.6) and (2.7) is the following: If

$$\prod_\nu F(x'_\nu, \lambda_1, \cdots, \lambda_s) = K(y'_1, \cdots, y'_p, \lambda_1, \cdots, \lambda_s) \cdot l'(x'_1, \cdots, x'_n),$$

where the Borel-measurable function y'_i is the same function of $x'_1, \cdots, x'_n$ that y_i is of $x_1, \cdots, x_n$, and where K and l' are real single-valued measurable functions of their indicated variables, then what is the relation between $\mathcal{D}(\mathbf{x}_\nu)$ and $\mathcal{D}(\mathbf{x}'_\nu)$?

The answer to that question is contained, for the absolutely continuous

* Sufficient statistics were first defined by R. A. Fisher; see [9], p. 319. Also see [26], where certain definitions of sufficient statistics are compared.

statistics and probability densities, in the corollary to the following derivation of the form which $\mathcal{D}(x_\nu)$ must have in order that a system of sufficient statistics exist.*

THEOREM 3. *Let $f(x_\nu, \theta_1, \cdots, \theta_q)$ be a non-vanishing continuous function of $x_\nu, \theta_1, \cdots, \theta_q$ on an open subset† V of‡ $R^1 \times R^q$, and let f be an absolutely continuous function of x_ν on $V(\theta_1^0, \cdots, \theta_q^0)$ for every set of fixed values, $\theta_1^0, \cdots, \theta_q^0$, of $\theta_1, \cdots, \theta_q$ in Θ. Set $\mathcal{D}(\mathbf{x}_\nu) = f(x_\nu, \theta_1, \cdots, \theta_q)$, and assume $\mathbf{x}_1, \cdots, \mathbf{x}_n$ to be independently distributed. Let (2.9) be satisfied, it now being assumed that y_i is an absolutely continuous function of x_ν on an open subset B of R' containing A, for fixed values of $x_1, \cdots, x_{\nu-1}, x_{\nu+1}, \cdots, x_n$. If we assume that the jacobian, $J(\delta_1, \cdots, \delta_p)$, where*

$$J(\delta_1, \cdots, \delta_p) = \frac{\partial(y_1, \cdots, y_p)}{\partial(x_{\delta_1}, \cdots, x_{\delta_p})},$$

does not vanish on $(B-Z)^n$, where Z is a closed set of measure zero, for any of the C_p^n possible selections $\delta_1, \cdots, \delta_p$ of p of the positive integers $1, \cdots, n$, then there exists an open subset V' of V such that the measure of $V - V'$ is zero, and $J(\delta_1, \cdots, \delta_p)$ does not vanish on any of the set $[V'(\theta_1, \cdots, \theta_q)]^n$ for all possible sets of values of $\delta_1, \cdots, \delta_p$. Furthermore, at each point $x_\nu^0, \theta_1^0, \cdots, \theta_p^0$ of V', there is neighborhood $N = \omega \leqq r$, where the definitions of ω and r are

$$\omega: \quad |x_\nu - x_\nu^0| < h, \qquad r: \quad |\theta_i - \theta_i^0| < h_i, \qquad i = 1, \cdots, q,$$

in which there exist real single-valued absolutely continuous non-constant functions $\alpha_j(x_\nu)$ and $\gamma(x_\nu)$ of x_ν and real single-valued continuous functions $\beta_j(\theta_1, \cdots, \theta_q)$ and $\beta(\theta_1, \cdots, \theta_q)$ of $\theta_1, \cdots, \theta_q$, such that in N

$$f(x_\nu, \theta_1, \cdots, \theta_q) = \exp\left[\sum_j \alpha_j(x_\nu)\beta_j(\theta_1, \cdots, \theta_q) + \gamma(x_\nu) + \beta(\theta_1, \cdots, \theta_q)\right]. \tag{2.10}$$

Proof. It is easy to see that if we define V' to be the intersection of V and $(B-Z)\times\Theta$, then V' is open, and the measure of $V - V'$ is zero. Now let the

* The result (2.10) has been obtained by Koopman and Darmois. See [23], p. 402 and [8], p. 1265. I have not been able to obtain the corollary to Theorem 3 from Koopman's proof, which depends on a definition of sufficient statistics that includes the definition based on (2.9). It may be remarked that Koopman needs only the continuity of the functions y_i whereas absolute continuity is assumed in Theorem 3. The proofs of Theorems 3 and 4 are generalizations of proofs for $p=q=1$ in (2.9) given by Doob in an unpublished seminar lecture at Columbia University. R. A. Fisher has derived Theorems 3 and 4 in cases which are specializations of those here treated. See [15].

† The totality of elements in the sets $V(\theta_1, \cdots, \theta_q)$ will be denoted by A.

‡ The values of x_ν are elements of R^1, and the values of $\theta_1, \cdots, \theta_q$ are elements of an open subset Θ of R^q.

neighborhood N be defined as in the hypotheses. Then the difference,

$$\sum_{\nu} [\log f(x_\nu, \theta_1, \cdots, \theta_q) - \log f(x_\nu, \theta_1', \cdots, \theta_q')],$$

is a real single-valued continuous function of the variables $x_1, \cdots, x_n$, $\theta_1, \cdots, \theta_q, \theta_1', \cdots, \theta_q'$, on $\omega^n \times r^2$ and is absolutely continuous in x_ν for fixed values of the other variables. Hence the equation

$$\frac{\partial}{\partial x_\nu} [\log f(x_\nu, \theta_1, \cdots, \theta_q) - \log f(x_\nu, \theta_1', \cdots, \theta_q')]$$

$$= \sum_i \frac{\partial y_i}{\partial x_\nu} \cdot \frac{\partial}{\partial y_i} [\log k(y_1, \cdots, y_p, \theta_1, \cdots, \theta_q) - \log k(y_1, \cdots, y_p, \theta_1', \cdots, \theta_q')]$$

is valid throughout $\omega^n \times r^2$. Letting $\delta_j = j+1$ in $J(\delta_1, \cdots, \delta_p)$, which involves no loss of generality, we see that the equation

$$\frac{\partial}{\partial y_i} [\log k(y_1, \cdots, y_p, \theta_1, \cdots, \theta_q) - \log k(y_1, \cdots, y_p, \theta_1', \cdots, \theta_q')]$$

$$= \sum_{\mu=2}^{p+1} a_{\mu i} \frac{\partial}{\partial x_\mu} [\log f(x_\mu, \theta_1, \cdots, \theta_q) - \log f(x_\mu, \theta_1', \cdots, \theta_q')],$$

where the $a_{\mu i}$ is the cofactor of $\partial y_i / \partial x_\mu$ in $J(2, \cdots, p+1)$ divided by $J(2, \cdots, p+1)$, is valid throughout $\omega^n \times r^2$. Upon assigning fixed values, $x_2^0, \cdots, x_n^0, \theta_1'^0, \cdots, \theta_q'^0$ from $\omega^{n-1} \times r$ to $x_2, \cdots, x_n, \theta_1', \cdots, \theta_q'$, we find that the equation

$$\frac{\partial}{\partial x_1} \log f(x_1, \theta_1, \cdots, \theta_q) = h(x_1) + \sum_j \bar{\alpha}_j(x_1) \beta_j(\theta_1, \cdots, \theta_q), \tag{2.11}$$

where

$$\begin{aligned} h(x_1) &= f(x_1, \theta_1'^0, \cdots, \theta_q'^0), \\ \bar{\alpha}_j(x_1) &= \sum_i \frac{\partial y_i}{\partial x_1} a_{j+1,i}, \\ \beta_j(\theta_1, \cdots, \theta_q) &= \frac{\partial}{\partial x_{j+1}} [\log f(x_{j+1}, \theta_1, \cdots, \theta_q) - \log f(x_{j+1}, \theta_1'^0, \cdots, \theta_q'^0)], \\ x_{j+1} &= x_{j+1}^0, \end{aligned} \tag{2.12}$$

is valid throughout $\omega \times r$. Then (2.10) may be obtained by integration of (2.11).

The absolute continuity and continuity of the functions may be easily proved. Obviously, if (2.10) is valid for all possible values of x_ν and $\theta_1, \cdots, \theta_q$, then the statistics, $\sum_\nu \alpha_j(\mathbf{x}_\nu)$, are a sufficient system of statistics with respect to the estimation of $\theta_1, \cdots, \theta_q$.

It is clear that a theorem similar to Theorem 3 may be proved for multivariate distributions.

The answer to the question of the relation of densities determining the same sufficient statistics is contained in the following corollary:

COROLLARY. *Let*

$$\mathfrak{D}(\mathbf{x}_\nu) = f(x_\nu, \theta_1, \cdots, \theta_q),$$
$$\mathfrak{D}(\mathbf{x}'_\nu) = F(x'_\nu, \lambda_1, \cdots, \lambda_s),$$

and let f and F not vanish except on a set of values of x of measure zero for all values of $\theta_1, \cdots, \theta_q$ and $\lambda_1, \cdots, \lambda_s$ considered. Assume that $\mathfrak{D}(x_\nu)$ and $\mathfrak{D}(x'_\nu)$ satisfy the conditions of Theorem 3, the statistics $y'_1, \cdots, y'_p$ being the same functions of $x'_1, \cdots, x'_n$ that $y_1, \cdots, y_p$ are of $x_1, \cdots, x_n$. Then in N and N', when N' is a neighborhood of values of $x', \lambda_1, \cdots, \lambda_s$ satisfying the conditions of Theorem 3, the x' values of N' being those of ω, the equations (2.10) *and*

$$F(x'_\nu, \lambda_1, \cdots, \lambda_s) = \exp\left[\sum_j \alpha_j(x'_\nu)\beta'_j(\lambda_1, \cdots, \lambda_s) + \beta'(\lambda_1, \cdots, \lambda_s) + \gamma'(x'_\nu)\right] \tag{2.13}$$

are valid; hence $\mathfrak{D}(\mathbf{x}_\nu)$ and $\mathfrak{D}(\mathbf{x}'_\nu)$ differ only in the functions γ, β, β_j and $\gamma', \beta', \beta'_j$. Furthermore if

$$l'(x_1, \cdots, x_n) \equiv l(x_1, \cdots, x_n),$$

then $\gamma(x) \equiv \gamma'(x)$, any constant difference being eliminated by means of β and β'.

The proof is omitted. It depends on (2.12).

We now assume (2.9) to be valid throughout $R^1 \times \Theta$ and find the joint probability density of the sufficient statistics $\phi_1, \cdots, \phi_p$ where

$$\phi_j = \frac{1}{n}\sum_\nu \alpha_j(x_\nu).$$

Integrating f with respect to x_ν we note that $\beta(\theta_1, \cdots, \theta_q)$ is a function $\delta(\beta_1, \cdots, \beta_p)$ of the values of $\beta_1(\theta_1, \cdots, \theta_q), \cdots, \beta_p(\theta_1, \cdots, \theta_q)$, defined by the equation

$$\exp\left[\delta(\beta_1, \cdots, \beta_p)\right] = \int_{-\infty}^{\infty} \exp\left[\sum_j \alpha_j(x)\beta_j(\theta_1, \cdots, \theta_q) + \gamma(x)\right] dx.$$

THEOREM 4. *Let* $\mathcal{D}(\mathbf{x}_\nu)$ *be given by* (2.10) *for all* x_ν *except on a set where* $\mathcal{D}(\mathbf{x}_\nu)=0$. *Then*

$$\mathcal{D}(\phi_1, \cdots, \phi_p) = (2\pi)^{-p} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left[i \sum_j t_j \phi_j - n\delta(\beta_1, \cdots, \beta_p) + n\delta\left(\beta_1 + \frac{it_1}{n}, \cdots, \beta_p + \frac{it_p}{n}\right)\right] dt_1 \cdots dt_p.$$

Proof. Since, by Lemma 1,

$$\mathcal{E}\left\{\exp\left[i \sum_j t_j \phi_j\right]\right\} = \mathcal{E}\left\{\exp\left[i \sum_j t_j \frac{\sum_\nu \alpha_j(\mathbf{x}_\nu)}{n}\right]\right\},$$

it follows that

$$\mathcal{E}\left\{\exp\left[i \sum_j t_j \phi_j\right]\right\} = \exp\left[- n\delta(\beta_1, \cdots, \beta_p) + n\delta\left(\beta_1 + \frac{it_1}{n}, \cdots, \beta_p + \frac{it_p}{n}\right)\right].$$

Then the proof is completed by using the Fourier transform theorem.*

One of the uses of Theorem 4 is to find the distribution of statistics which are functions of sufficient statistics, and which may, in fact, be an equivalent set of sufficient statistics.† It is easy to see that

$$\mathcal{E}\{\alpha_j(\mathbf{x})\} = \frac{\partial}{\partial \beta_j} \delta(\beta_1, \cdots, \beta_p),$$

and that

$$\sigma_{jk} = \mathcal{E}\{[\alpha_j(\mathbf{x}) - \mathcal{E}\{\alpha_j(\mathbf{x})\}][\alpha_k(\mathbf{x}) - \mathcal{E}\{\alpha_k(\mathbf{x})\}]\} = \frac{\partial^2}{\partial \beta_j \partial \beta_k} \delta(\beta_1, \cdots, \beta_p).$$

The purpose of the following theorem is to make possible the simplification of our proofs of Theorems 6, 7, and 8.

THEOREM 5. *Let the real single-valued Borel-measurable functions,*

$$y_\nu = f_\nu(x_1, \cdots, x_n), \qquad n > 1, \tag{2.14}$$

be defined on R^n. *Let* $y_{p_{\gamma-1}+1}, \cdots, y_{p_\gamma}$ *depend only on* $x_1, \cdots, x_{p_\gamma}$, *and let* $y_{p_{\gamma-1}+1}, \cdots, y_{p_\gamma}$ *be a linear transformation, having unit determinant, of* $x_{p_{\gamma-1}+1}, \cdots, x_{p_\gamma}$ *for almost all sets of fixed values of* $x_1, \cdots, x_{p_{\gamma-1}}$, *where the*

* See, for example, [5], p. 191.

† Two sets of sufficient statistics may be called equivalent if both satisfy equations of the form (2.9).

transformation may depend on the fixed values of $x_1, \cdots, x_{p_{\gamma-1}}$ *and* $n = p_m$. *Then if* A *is a Borel-measurable set of values of* $y_1, \cdots, y_n$, *and* A^{-1} *is the set of values of* $x_1, \cdots, x_n$ *such that* $y_1, \cdots, y_n$ *is an element of* A, *it follows that* A^{-1} *is Borel-measurable and has the same measure as* A.

Proof. The proof is given for $m = 2$. We shall use p for p_1. It is well known that if A is Borel-measurable, then A^{-1} is Borel-measurable. Let us denote by $c_A(y_1, \cdots, y_n)$ the function which has value one if $y_1, \cdots, y_n$ is an element of A and which is zero otherwise. If we denote the measure of A by $\mu(A)$, it follows that $c_A(y_1, \cdots, y_n)$ is Borel-measurable if A is Borel-measurable, and

$$\mu(A) = \int \cdots \int_{R^n} C_A(y, \cdots, y_n) dy_1 \cdots dy_n.$$

Now it may readily be shown that if $c_A(y_1, \cdots, y_n)$ is Borel-measurable in R^n, and we assign fixed values, $y_1^0, \cdots, y_p^0$ to $y_1, \cdots, y_p$, then $c_A(y_1^0, \cdots, y_p^0, y_{p+1}, \cdots, y_n)$ is Borel-measurable in R^{n-p}.* From Fubini's theorem,† it follows that

$$\mu(A) = \int \cdots \int_{R^p} \Big[\int \cdots \int_{R^{n-p}} C_A(y, \cdots, y_p, y_{p+1}, \cdots, y_n) dy_{p+1} \cdots dy_n \Big] dy_1 \cdots dy_p. \tag{2.15}$$

In a precisely similar fashion, it may be shown that

$$\mu(A^{-1}) = \int \cdots \int_{R^p} \Big[\int \cdots \int_{R^{n-p}} C_{A^{-1}}(x, \cdots, x_p, x_{p+1}, \cdots, x_n) dx_{p+1} \cdots dx_n \Big] dx_1 \cdots dx_p.$$

Inasmuch as the linear transformation having unit determinant is measure preserving, it follows that if our fixed values of $x_1, \cdots, x_p$ are determined by means of the assigned fixed values of $y_1, \cdots, y_p$ and the linear transformation, then, except for a set of measure zero,

$$\int \cdots \int_{R^{n-p}} C_A(y_1, \cdots, y_p, y_{p+1}, \cdots, y_n) dy_{p+1} \cdots dy_n = \int \cdots \int_{R^{n-p}} C_{A^{-1}}(x_1, \cdots, x_p, x_{p+1}, \cdots, x_n) dx_{p+1} \cdots dx_n.$$

* See, for example, [31], p. 82

† See [31], p. 77.

Since the transformation of $x_1, \cdots, x_p$ into $y_1, \cdots, y_p$ is measure preserving, the proof is completed by transforming (2.15).

COROLLARY. *Let $\psi(y_1, \cdots, y_n)$ be a real single-valued measurable function of $y_1, \cdots, y_n$ on R^n. Then $\psi(y_1, \cdots, y_n)$ is a real single-valued measurable function $\phi(x_1, \cdots, x_n)$ of $x_1, \cdots, x_n$ on R^n when we substitute from* (2.14), *and*

$$\int \cdots \int_A \psi(y_1, \cdots, y_n) dy_1 \cdots dy_n = \int \cdots \int_{A^{-1}} \phi(x_1, \cdots, x_n) dx_1 \cdots dx_n.$$

Proof. If A has finite measure, we may use Lemma 1 to complete the proof, and if A is not of finite measure, we express A as a sum of non-overlapping sets of finite measure and then proceed as above.

More general results than Theorem 5 and its corollary may readily be obtained. It is only necessary to require that A and A^{-1} be in 1-1 correspondence, to replace the words "a linear transformation having unit determinant" by the words "absolutely continuous functions with non-vanishing jacobian," and to replace the words "and has the same measure as A" by the words "and has measure

$$\int \cdots \int_{A^{-1}} J_1 \cdots J_m dx_1 \cdots dx_n,$$

where

$$J_\gamma = \left| \frac{\partial y_i}{\partial x_j} \right|, \qquad i, j = p_{\gamma-1} + 1, \cdots, p_\gamma."$$

Similar theorems may be proved when R^n has a completely additive measure function defined on it,* and the functions (2.14) are measurable with respect to that measure function.

The theorems stated next are of constant service in the analysis of the normal multivariate distribution.

THEOREM 6. *Let† $p = p_1 + \cdots + p_m$, let*

$$\mathfrak{D}(\mathbf{x}_{11}, \cdots, \mathbf{x}_{pn}) = \prod_\gamma \mathfrak{D}(\mathbf{x}_{p'_{\gamma-1}+1,1}, \cdots, \mathbf{x}_{p'_\gamma,n}),$$

$$p'_0 = 0; \; p'_\gamma = p_1 + \cdots + p_\gamma,$$

and assume that a positive integer, $\gamma_0 < m$, exists such that for $\gamma > \gamma_0$

$$\mathfrak{D}(\mathbf{x}_{p'_{\gamma-1}+1,1}, \cdots, \mathbf{x}_{p'_\gamma,n}) = g_\gamma(a_{ij}),$$

---

* See, for example, [31], chap. 1.

† In this theorem we do not define p_γ to be $n_1 + \cdots + n_\gamma$.

where it is understood that g_γ is a function of all the a_{ij}, $(i, j = p_{\gamma-1}+1, \cdots, p'_\gamma)$ and $a_{ij} = \sum_\nu x_{i\nu}x_{j\nu}$. Let the real single-valued Borel-measurable functions

$$y_{i\nu} = f_{i\nu}(x_{11}, \cdots, x_{pn})$$

be defined on R^{pn}, and let the equations,

$$y_{i\nu} = \sum_{\mu=1}^{n} \phi_{\mu\nu}^{(\gamma)}(x_{11}, \cdots, x_{p'_{\gamma-1},n})x_{i\mu}, \qquad i = p'_{\gamma-1}+1, \cdots, p'_\gamma,$$

define a non-singular linear transformation of $x_{i1}, \cdots, x_{in}$ into $y_{i1}, \cdots, y_{in}$ for fixed values of $x_{11}, \cdots, x_{p'_{\gamma-1},n}$ and each value of i, the transformation being orthogonal if $\gamma > \gamma_0$.

Then, if

$$b_{ij} = \sum_\nu y_{i\nu}y_{j\nu},$$

it follows that

$$\mathcal{D}(y_{11}, \cdots, y_{pn}) = \mathcal{D}(y_{11}, \cdots, y_{p'_{\gamma_0}-1,n}) \cdot \prod_{\gamma=\gamma_0+1}^{m} \mathcal{D}(y_{p'_{\gamma-1}+1,1}, \cdots, y_{p'_\gamma,n}),$$

where

$$\mathcal{D}(y_{p'_{\gamma-1}+1,1}, \cdots, y_{p'_\gamma,n}) = g_\gamma(b_{ij})$$

and it is understood that g_γ is a function of all the b_{ij}, $(i, j = p'_{\gamma-1}+1, \cdots, p'_\gamma)$.

Proof. The functions $y_{i\nu}$ have been so defined that the corollary to Theorem 5 may be applied.

We come now to the generalizations of the Fisher-Cochran theorem.

THEOREM 7. *Let*

$$\mathcal{D}(\mathbf{x}_{1\nu}, \cdots, \mathbf{x}_{p\nu}) = N((x_\nu); (\sigma)),$$

and let

$$\mathcal{D}(\mathbf{x}_{11}, \cdots, \mathbf{x}_{pn}) = \prod_\nu \mathcal{D}(\mathbf{x}_{1\nu}, \cdots, \mathbf{x}_{p\nu}).$$

Let, for each value of i, j, and γ the real single-valued function

$$b_{ij\gamma} = \sum_{\mu,\nu=1}^{n} d_{\mu\nu}^{\gamma} x_{i\mu}x_{j\nu}, \tag{2.16}$$

be a bilinear form in $x_{i\nu}$ and $x_{j\nu}$ with constant coefficients, and let the matrix of $b_{ij\gamma}$ be denoted by (d^γ). Suppose that

$$\sum_\gamma b_{ij\gamma} = \sum_\nu x_{i\nu}x_{j\nu}. \tag{2.17}$$

Then, if n_γ be the rank of (d^γ), a necessary and sufficient condition that there exist linear functions $z_{i\nu}$ of $x_{i1}, \cdots, x_{in}$ such that $\mathcal{D}(z_{1\nu}, \cdots, z_{p\nu}) = N((z_\nu); (\sigma))$,

$$\mathcal{D}(z_{11}, \cdots, z_{p\nu}) = \prod_\nu \mathcal{D}(z_{1\nu}, \cdots, z_{p\nu})$$

and

$$\sum_{\nu=p_{\gamma-1}+1}^{p_\gamma} z_{i\nu}z_{j\nu} = b_{ij\gamma}$$

is

$$p_m = n. \tag{2.18}$$

Proof. The necessity of the condition (2.18) is obvious. In order to prove that (2.18) is sufficient, let $i=j=1$. Then by Theorem 1, there exist n linear functions (2.2) which have the desired properties. Consider the system of linear functions

$$z_{i\mu} = \sum_\nu c_{\mu\nu}x_{i\nu}, \qquad \mu = 1, \cdots, n, \tag{2.19}$$

the coefficients of which are those of the linear functions (2.2). The functions (2.19) are an orthogonal transformation of the chance variables $x_{i\nu}$ for fixed i and have the desired properties.

COROLLARY 1. *If*

$$n_\gamma \geqq p, \qquad \gamma = 1, \cdots, k;\ k \leqq m, \tag{2.20}$$

and if $n \geqq p_k$, then the chance variables

$$b_{ij\gamma}, z_{i\mu}, \qquad \gamma = 1, \cdots, k; \mu = p_\gamma + 1, \cdots, n, \tag{2.21}$$

have density

$$(2\pi)^{-p(p-1)k/4-(n-p_k)/2}\sigma^{-n/2}\left\{\prod_{\gamma=1}^{k}\prod_{i=1}^{p}[\Gamma((n_\gamma - i + 1)/2)]^{-1}\right\}$$
$$\cdot\left\{\prod_{\gamma=1}^{k}|b_{ij\gamma}|^{(n_\gamma-p+1)/2-1}\right\}\exp\left[-\frac{1}{2}\sum_{i,j}\sigma^{ij}\left\{\sum_{\gamma=1}^{k}b_{ij\gamma}+\sum_{\nu=p_k+1}^{n}z_{i\nu}z_{j\nu}\right\}\right],$$

where $|b_{ij\gamma}|$ is a p rowed determinant for each value of γ.

Proof. The condition (2.20) permits the use of Wishart's distribution.*

COROLLARY 2. *Let $\mathcal{D}(x_{11}, \cdots, x_{pn}) = \prod_\nu N((x_\nu); (\sigma))$, and let the bilinear forms (2.16) have the property (2.17). Then if the linear functions $z_{i\nu}$ exist for $i=1, \cdots, p$; $\nu=1, \cdots, p_{m-1}$, and if*

$$\mathcal{D}(z_{11}, \cdots, z_{p,p_{m-1}}) = \prod_{\nu=1}^{p_{m-i}} N((z_\nu); (\sigma)),$$

* See equation (2.8) of this paper, or Wishart [39], p. 38, or Wishart and Bartlett [4], p. 270.

it follows that

$$n_m = n - p_{m-1},$$

that the complete set of variables (2.19) *exists, and that*

$$\mathcal{D}(z_{11}, \cdots, z_{pn}) = \prod_{\nu} N((z_\nu); (\sigma)).$$

Proof. See [6], p. 181.*

In Theorem 7 a direct generalization of the Fisher-Cochran theorem was considered. In Theorem 8, the generalization is one which permits the use of more general quadratic forms than could be examined by means of Theorem 7. However, transformations of the chance variable (2.21) such as that of Wilks [34], p. 484, equation (30), will provide another derivation, by the moment method, of many of the distributions which we shall obtain, in §3, by means of Theorems 8 and 9.

THEOREM 8. *Let the* $n_1+n_2+\cdots+n_p$ *chance variables* $x_{i\nu}$, $(i=1, \cdots, p; \nu=1, \cdots, n_i)$, *have probability density* $\prod_{i=1}^{p}\prod_{\nu=1}^{n_i} N(x_{i\nu}; 1)$. *Let the real single-valued Borel-measurable function*

$$q_{ij} = \sum_{\mu,\nu=1}^{n_i} \omega_{\mu\nu}^{ij}(x_{11}, \cdots, x_{i-1,n_{i-1}}) x_{i\mu} x_{i\nu} \tag{2.22}$$

of $x_{11}, \cdots, x_{in}$, *defined on* $R^{n_1+\cdots+n_p}$, *be, for each pair of values of* i *and* j, $(i=1, \cdots, p;\ j=1, \cdots, m_i)$, *and for almost all sets of fixed values of* $x_{11}, \cdots, x_{i-1,n_{i-1}}$, *a quadratic form in* $x_{i1}, \cdots, x_{in_i}$ *of rank* n_{ij}, *and let*

$$\sum_{j=1}^{m_i} q_{ij} = \sum_{\nu=1}^{n_i} x_{i\nu}^2, \qquad i = 1, \cdots, p. \tag{2.23}$$

Then a necessary and sufficient condition that

$$\mathcal{D}(q_{11}, \cdots, q_{pk_p}) = \prod_{i=1}^{p}\prod_{j=1}^{m_i} G(q_{ij}; n_{ij}, 1) \tag{2.24}$$

is

$$\sum_{j=1}^{m_i} n_{ij} = n_i. \tag{2.25}$$

* It is not difficult to generalize Theorems II and III of Cochran ([16], pp. 179, 181) by following the same procedure that has sufficed for the generalization of Cochran's Theorem II. It may be remarked that a necessary and sufficient condition for the satisfaction of equation (3.1) of Wilks ([37], p. 326) is provided by the generalization of Cochran's Theorem III. The generalizations of Cochran's Theorems I and II are also very useful in the moment approach to the theory of multivariate statistical analysis.

Proof. *Necessity.* The additive property of the χ^2 distribution suffices; for, $\sum_{\nu=1}^{n_i} x_{i\nu}^2$ has the χ^2 distribution with n_i degrees of freedom, and from (2.22) and (2.23), $\sum_{\nu=1}^{n_i} x_{i\nu}^2$ has the χ^2 distribution with $n_{i1}+n_{i2}+\cdots+n_{im_i}$ degrees of freedom.

Sufficiency. We obtain the analogue of Theorem 1 for the hypotheses of Theorem 8. For each value of i, assigning fixed values to $x_{11}, \cdots, x_{i-1,n_{i-1}}$, we obtain from Theorem 1 n_i real linear functions

$$z_{i\nu} = \sum_{\mu=1}^{n_i} c_{\mu\nu}^{i} x_{i\mu}, \tag{2.26}$$

which are such that the corresponding chance variables have density

$$\prod_{\nu=1}^{n_i} N(z_{i\nu}; 1);$$

and for the fixed values of $x_{11}, \cdots, x_{i-1,n_{i-1}}$

$$\sum_{\nu} z_{i\nu}^2 = q_{ij}, \qquad j = 1, \cdots, m_i,$$

where ν runs from $n_{i1}+\cdots+n_{i,j-1}+1$ through $n_{i1}+\cdots+n_{ij}$.

The coefficients $c_{\mu\nu}^{i}$, of the linear forms (2.26) are real single-valued Borel-measurable functions of the coefficients of q_{ij} for fixed values of $x_{11}, \cdots, x_{i-1,n_{i-1}}$. Let $c_{\mu\nu}^{i}(\omega)$ be the same function of the functions

$$\omega_{\mu\nu}^{ij}(x_{11}, \cdots, x_{i-1,n_{i-1}}), \qquad \mu, \nu = 1, \cdots, n_i,$$

that $c_{\mu\nu}^{i}$ is of the coefficients of the quadratic form having constant coefficients. Since Borel-measurable functions of Borel-measurable functions are Borel-measurable, it follows that the functions

$$\begin{aligned} y_{i\nu} &= z_{i\nu}(x_{11}, \cdots, x_{in_i}) \\ &= \sum_{\mu=1}^{n_i} c_{\mu\nu}^{i}(\omega) x_{i\mu} \end{aligned}$$

are Borel-measurable functions of $x_{11}, \cdots, x_{in_i}$. For almost all sets of fixed values of $x_{11}, \cdots, x_{i-1,n_{i-1}}$, it is clear that $y_{i1}, \cdots, y_{in_i}$ is a linear orthogonal transformation of $x_{i1}, \cdots, x_{in_i}$. Hence the hypotheses of Theorem 5 are satisfied. Furthermore, the functions $y_{i\nu}$ have been so defined that

$$\sum_{\nu} y_{i\nu}^2 = q_{ij}, \qquad j = 1, \cdots, m_i,$$

where ν runs from $n_{i1}+\cdots+n_{i,j-1}+1$ through $n_{i1}+\cdots+n_{ij}$. Then by the corollary to Theorem 5, it follows that

$$\mathcal{D}(y_{11}, \cdots, y_{pn_p}) = \prod_{i=1}^{p} \prod_{\nu=1}^{n_i} N(y_{i\nu}; 1),$$

and the proof is completed by using the additive property of the χ^2 distribution.

THEOREM 9. *Let* $\mathcal{D}(x_{11}, \cdots, x_{pn}) = \prod_\nu N((x_\nu); (\sigma))$. *Let* $n_1 = \cdots = n_p = n$ *in Theorem* 8, *and let the functions* (2.22) *satisfy* (2.23). *If, for some* i *and* j, $\sigma^{ij} \neq 0$, *let* $\psi_{ij}(q_{11}, \cdots, q_{pm_p})$ *be a real-valued Borel-measurable function of its indicated variables such that*

$$\psi_{ij}(q_{11}, \cdots, q_{pm_p}) = \sum_\nu x_{i\nu} x_{j\nu}.$$

Then a necessary and sufficient condition that

$$\mathcal{D}(q_{11}, \cdots, q_{pm_p}) = \left\{ \prod_i \prod_{\nu=1}^{m_i} [\Gamma(n_{i\nu}/2)]^{-1} q_{i\nu}^{n_{i\nu}/2-1} \right\} \cdot 2^{-pn/2} \sigma^{-n/2} \exp\left[-\frac{1}{2} \sum_{i,j} \sigma^{ij} \psi_{ij}(q_{11}, \cdots, q_{pm_p}) \right] \tag{2.27}$$

is

$$\sum_{\nu=1}^{m_i} n_{i\nu} = n. \tag{2.28}$$

Proof. *Necessity.* The necessity of the condition (2.28) may be demonstrated as in the proof of Theorem 8.

Sufficiency. If, in Theorem 2, we let $l = l' \equiv 1$,

$$K(q_{11}, \cdots, q_{pm_p}) = (2\pi)^{-np/2} \sigma^{-n/2} \exp\left[-\frac{1}{2} \sum_{i,j} \sigma^{ij} \psi_{ij}(q_{11}, \cdots, q_{pm_p}) \right],$$

and

$$k(q_{11}, \cdots, q_{pm_p}) = (2\pi)^{-np/2} \exp\left[-\frac{1}{2} \sum_i \psi_{ii}(q_{11}, \cdots, q_{pm_p}) \right],$$

then the functions (2.22) are such that (2.6) and (2.7) are true. Hence, the desired result, (2.27), follows from the straightforward application of Theorems 2 and 8.

It is apparent that many distributions of functions of chance variables having a normal multivariate distribution may be derived by using Theorem 9. Some of these will be considered in §3.

3. **Vector variates.** In this section we apply the methods developed in §2

to various problems in the generalized analysis of variance, relations between sets of variables, and generalized periodogram analysis.

In order to do this with a minimum of duplication, two theorems in the theory of quadratic forms are first stated. These theorems and their corollaries are useful in the theory of univariate statistical analysis. In Theorems 12 and 13, the previous theorems are used to obtain joint distributions of functions which are quadratic forms in certain variables for fixed values of other variables. These latter distributions are then transformed in order to derive distributions which are useful in the theory of vector variates. The notation and terminology have already been given in the first section of this paper.

Let

$$\sigma_i = \begin{vmatrix} \sigma_{11} & \sigma_{12} \cdots \sigma_{1i} \\ \sigma_{21} & \sigma_{22} \cdots \sigma_{2i} \\ \cdot & \cdot \quad \cdots \quad \cdot \\ \sigma_{i1} & \sigma_{i2} \cdots \sigma_{ii} \end{vmatrix}, \qquad \sigma_p = \sigma,$$

let $\sigma^i = \sigma_i/\sigma_{i-1}$, and let σ^{jki} be the cofactor of σ_{jk} in σ_i divided by σ_i, $(j, k = 1, \cdots, i)$, $\sigma^{jkp} = \sigma^{jk}$.

Set

$$a_{ij} = \sum_{\nu} x_{i\nu} x_{j\nu},$$

and assume that a_i, a^i, and a^{jki} are the same functions of the a_{ij} that σ_i, σ^i, and σ^{jki} are of the σ_{ij}. We shall refer to any determinant a_i as a generalized sum of squares since

$$a_i = \sum_{\mu} (\Delta_\mu)^2,$$

where

$$\Delta_\mu = \begin{vmatrix} x_{1\mu_1} & x_{1\mu_2} \cdots x_{1\mu_i} \\ x_{2\mu_1} & x_{2\mu_2} \cdots x_{2\mu_i} \\ \cdot & \cdot \quad \cdots \quad \cdot \\ x_{i\mu_1} & x_{i\mu_2} \cdots x_{i\mu_i} \end{vmatrix}, \tag{3.1}$$

and the summation is for all the C_i^n distinct selections of i integers $\mu_1, \mu_2, \cdots, \mu_i$ from $1, \cdots, n$.

We remark that if β_{ig} is the regression coefficient of x_g in the "true" regression function of x_i on $x_1, \cdots, x_{i-1}$, where $x_1, \cdots, x_p$, $(p \geqq i)$, are assumed to have a normal multivariate distribution, then

$$\beta_{ig} = -\frac{\sigma^{igi}}{\sigma^{iii}}.$$

Similarly, if b_{ig} is the regression coefficient of x_g in the sample regression function of x_i on $x_1, \cdots, x_{i-1}$, then

$$b_{ig} = -\frac{a^{igi}}{a^{iii}}.$$

In the following theorem it is shown that the coefficients of a transformation by which a definite quadratic form may be reduced to a sum of squares are the regression coefficients. It is necessary to emphasize that although much of this material is known, yet no complete treatment seems to exist.

THEOREM 10. *There exist p linear forms*

$$y_i = x_i - \sum_g \beta_{ig} x_g \tag{3.2}$$

such that

$$q = \sum_i y_i^2/\sigma^i.$$

The proof is omitted. It is noted that the transformation (3.2) has unit determinant.

COROLLARY 1. *Let*

$$\mathcal{D}(\mathbf{x}_{1\nu}, \cdots, \mathbf{x}_{p\nu}) = N((x_\nu); (\sigma)),$$

and let

$$\mathcal{D}(\mathbf{x}_{11}, \cdots, \mathbf{x}_{pn}) = \prod_\nu \mathcal{D}(\mathbf{x}_{1\nu}, \cdots, \mathbf{x}_{p\nu}).$$

Then if $y_{i\nu} = x_{i\nu} - \sum_g \beta_{ig} x_{g\nu}$, *it follows that*

$$\mathcal{D}(\mathbf{y}_{i\nu}) = N(y_{i\nu}; \sigma^i)$$

and

$$\mathcal{D}(\mathbf{y}_{11}, \cdots, \mathbf{y}_{pn}) = \prod_\nu \prod_i \mathcal{D}(\mathbf{y}_{i\nu}).$$

The proof is omitted.

In Theorem 10 and Corollary 1 it was shown that variables which have a joint normal distribution may be expressed in terms of variables which are distributed independently in normal distributions. In the following corollaries it will be shown that certain functions of the dependent variables are invariant under the transformation into independent variables and hence have the distributions of the same functions of independent variables.

COROLLARY 2. *Let a_i' be the generalized sum of squares with the elements*

$$a'_{jk} = \sum_{\nu} y_{j\nu}y_{k\nu}, \qquad j, k = 1, \cdots, i.$$

Then

$$a_i' = a_i. \tag{3.3}$$

Proof. Equation (3.3) is obvious when a_i and a_i' are written as sums of squares of determinants as in (3.1).

COROLLARY 3. *If $a'^i = a_i'/a_{i-1}'$,* then, except for a set of measure zero,*

$$a'^i = a'_{ii} - \sum_{\varrho} w^2_{i\varrho}, \tag{3.4}$$

where

$$w_{i\varrho} = \sum_{\nu} p_{\varrho\nu}y_{i\nu} \tag{3.5}$$

and the coefficients

$$p_{\varrho\nu} = (a'^{\varrho})^{-1/2}\left(y_{\varrho\nu} - \sum_{j=1}^{\varrho-1} b'_{\varrho j}y_{j\nu}\right) \tag{3.6}$$

are such that

$$\sum_{\nu} p_{\varrho\nu}p_{k\nu} = \delta_{\varrho k}. \tag{3.7}$$

Proof. It follows from the definition of a'^i that†

$$a'^i = a'_{ii} - \sum_{\varrho,h} a'_{i\varrho}a'_{ih}a'^{\varrho h i-1}. \tag{3.8}$$

From Theorem 1 and (3.8), it follows that

$$a'_{ii} - a'^i = \sum_{\varrho} w^2_{i\varrho},$$

for almost all sets of fixed values of $y_{11}, \cdots, y_{i-1,n}$, where‡

$$\begin{aligned} w_{i\varrho} &= (a'^{\varrho})^{-1/2}\left(a'_{i\varrho} - \sum_{k=1}^{\varrho-1} b'_{\varrho k}a'_{ik}\right) \\ &= (a'^{\varrho})^{1/2}b'_{i\cdot\varrho}. \end{aligned} \tag{3.9}$$

---

* The functions a'^{jki} and $b'_{i\varrho}$ are to be defined as are a^{jki} and $b_{i\varrho}$ but with a'_{jk} replacing a_{jk}, $(j, k=1, \cdots, i)$. From Corollary 2 it follows that $a'^i = a^i$.

† It is, of course, well known that two alternative expressions for a'^i are $a'^i = a_{ii} - \sum_{\varrho} a_{i\varrho}b'_{i\varrho}$ and $a'^i = a_{ii} - \sum_{\varrho,h} b'_{i\varrho}b'_{ih}a'_{\varrho h}$.

‡ The function $b_{i\cdot\varrho}$ is the coefficient of x_ϱ in the sample regression equation of x_i on $x_1, \cdots, x_\varrho$, and $b'_{i\cdot\varrho}$ is the same function of the $y_{i\nu}$.

Hence, equations (3.4), (3.5), and (3.6) are true. Equation (3.7) may be verified by direct computation, or by employing the known properties of deviations from sample regression functions to avoid that computation.

COROLLARY 4. *The function w_{ig} may be written as*

$$w_{ig} = (a^{i})^{1/2}\left(b_{i\cdot g} - \sum_{k=g+1}^{i-1} b_{k\cdot g}\beta_{ik} - \beta_{ig}\right). \tag{3.10}$$

Proof. From (3.9) it follows that

$$w_{ig} = (a'^{g})^{-1/2}(a'_{g-1})^{-1}\begin{vmatrix} a'_{11} & \cdots & a'_{1g} \\ \cdot & \cdots & \cdot \\ a'_{g-1,1} & \cdots & a'_{g-1,g} \\ a'_{i1} & \cdots & a'_{ig} \end{vmatrix}.$$

Now it may be shown that if

$$K'_{qr} = \begin{vmatrix} a'_{11} & \cdots & a'_{1,g-1} & a'_{1q} \\ \cdot & \cdots & \cdot & \cdot \\ a'_{g-1,1} & \cdots & a'_{g-1,g-1} & a'_{g-1,q} \\ a'_{r1} & \cdots & a'_{r,g-1} & a'_{r,q} \end{vmatrix},$$

where q and r are positive integers such that $g \leqq q \leqq i$; $g \leqq r \leqq i$, then

$$K'_{qr} = \sum_{\mu=g}^{q}\sum_{\nu=g}^{r} \beta_{q\mu}\beta_{r\nu}K_{\mu\nu},$$

where K_{qr} is the same function of the a_{ij} that K'_{qr} is of the a'_{ij}. Then (3.10) is a consequence of the fact that

$$b_{kg} = \frac{K_{kg}}{a_g}, \qquad k = g, \cdots, i.$$

Although Theorem 10 and Corollaries 1–4 are more than sufficient for the consideration of the distribution of the generalized sum of squares, they do not yield the distribution of the generalized correlation ratio and related statistics. It is for this reason that the following definitions, theorem, and corollaries are introduced.

Let

$$a_{ij\gamma} = \sum_{\nu} x_{i\nu}x_{j\nu}, \qquad n_0 = 0;\ n_\gamma \geqq p;\ n_1 + \cdots + n_m \leqq n,$$

where ν runs from $n_1 + \cdots + n_{\gamma-1} + 1$ through $n_1 + \cdots + n_\gamma$, and let $a_{i(\gamma)}$ and $a^{i}_{(\gamma)}$, be defined as are a_i and a^i but with $a_{kj\gamma}$ replacing a_{kj}. Let $a'_{ij\gamma}$, $a'_{i(\gamma)}$,

and $a_{(\gamma)}^{\prime t}$ be similar functions of the $y_{i\nu}$. Then there exist linear functions $w_{ig(\gamma)}$ such that Corollary 3 is true for each value of γ.

Let s be a positive integer not greater than p, and let

$$\tau_i = \begin{vmatrix} \sigma_{ii} & \sigma_{i,s+1} & \cdots & \sigma_{ip} \\ \sigma_{s+1,i} & \sigma_{s+1,s+1} & \cdots & \sigma_{s+1,p} \\ \cdot & \cdot & \cdots & \cdot \\ \sigma_{pi} & \sigma_{p,s+1} & \cdots & \sigma_{pp} \end{vmatrix}, \qquad i = 1, \cdots, s.$$

Let τ^{jki} be the cofactor of σ_{jk} in τ_i divided by τ_i, $(j, k = i, s+1, \cdots, p; i = 1, \cdots, s)$. Let τ_0 be the cofactor of σ_{ii} in τ_i, and let $\tau_0{}^{jk}$ be the cofactor of σ_{jk} in τ_0 divided by τ_0. Let d be the same function of the a_{ij} that τ_0 is of the σ_{ij}.

THEOREM 11. *There exist p linear forms, for a preassigned value of ν,*

$$(3.11) \qquad \begin{aligned} z_{i\nu} &= x_{i\nu} + \sum_{j=s+1}^{p} \frac{\tau^{iji}}{\tau^{iii}} x_{j\nu}, & i &= 1, \cdots, s, \\ z_{i\nu} &= x_{i\nu}, & i &= s+1, \cdots, p, \end{aligned}$$

such that

$$q_\nu = \sum_{i,j=1}^{s} \sigma^{ij} z_{i\nu} z_{j\nu} + \sum_{i,j=s+1}^{p} \tau_0{}^{ij} z_{i\nu} z_{j\nu}.$$

The proof is omitted. It is noted that the transformation (2.36) has unit determinant.

R. A. Fisher* has evaluated the difference between the sum of squares of the deviations of y_i from the sample regression function in $y_1, \cdots, y_{i-1}$, and the sum of squares of the deviations of y_i from the sample regression function in $y_1, \cdots, y_k$, $(k < i-1)$. In order to state his result in terms of the functions which are generally computed in least squares solutions, we define the $i-k-1$ rowed determinant

$$\alpha' = |\, a'^{jt,i-1} \,|, \qquad j, t = k+1, \cdots, i-1,$$

and let α'_{jt} be the cofactor of $a'^{jt,i-1}$ in α' divided by α'. Then we may state Fisher's result as the following corollary:

COROLLARY 1. *If $a'^{i(k)}$ is the sum of squares of the deviations of y_i from its sample regression function in $y_1, \cdots, y_k$, then the difference, $a'^{i(k)} - a'^{i}$, is, for almost all sets of fixed values of $y_{11}, \cdots, y_{i-1,n}$, a positive definite quadratic form in $y_{i1}, \cdots, y_{in}$ of rank $i-k-1$:*

$$\sum_{j,t=k+1}^{i-1} b'_{ij} b'_{it} \alpha'_{jt}.$$

* See [16], §29.1.

The proof is omitted. A simple proof may be given by means of Theorem 11 and the Jacobi ratio theorem* in determinants.

We now use Corollary 1 to determine the change in the rank of a^i induced by the omission of several observations.

Let us define, for each value of t, $(t=1, \cdots, n-m)$, a variable which assumes only the values zero and one,

$$y_{i+t,\nu} = \delta_{\nu,n_1+t}.$$

Then $a'_{(1)}$ is the sum of squares of the deviations of y_i from its sample regression function in $y_1, \cdots, y_{i-1}, y_{i+1}, \cdots, y_{i+n-n_1}$.

COROLLARY 2. *The difference*

$$a'^i - a'^i_{(1)}$$

is, for fixed i and almost all sets of fixed values of $y_{11}, \cdots, y_{i-1,n}$, a quadratic form in $y_{i1}, \cdots, y_{in}$ of rank $n-n_1$.†

The proof is omitted.

The following corollary is useful in determining whether or not a set of regression coefficients differ significantly from one another.

Let a_i'' be the generalized sum of squares, the elements of which are $\sum_\gamma a_{jk\gamma}$, $(j, k=1, \cdots, i)$, and let $v_i = a''^i - a'^i_{(1)} - \cdots - a'^i_{(m)}$.

COROLLARY 3. *For any fixed value of i, and for almost all sets of fixed values of $y_{11}, \cdots, y_{i-1,n}$, the function v_i is a quadratic form in $y_{i1}, \cdots, y_{in}$ of rank $(m-1)(i-1)$ and is equal to*

$$\sum_\gamma \sum_{g,h} a_{gh\gamma}(b_{ig} - b_{ig}^{(\gamma)})(b_{ih} - b_{ih}^{(\gamma)}).\ddagger \tag{3.12}$$

The proof is omitted.§

Let $n' = n_1 + \cdots + n_m$, and let $u_i = a'^i - a''^i$, where $a''^i = a_i''/a_{i-1}''$. Then we can state the following theorem:

THEOREM 12. *Let*

$$\mathfrak{D}(\mathbf{x}_{1\nu}, \cdots, \mathbf{x}_{p\nu}) = N((x_\nu); (\sigma)),$$

and let

---

* Turnbull [32], p. 77.

† The formula obtained by substituting in (3.11) does not provide a simple evaluation of the difference.

‡ The functions $b_{ig}^{(\gamma)}$ are the regression coefficients in the γth set of observations; in other words, $b_{ig}^{(\gamma)}$ is defined as is b_{ig} but with $a_{ki\gamma}$ replacing a_{kj}.

§ Certain special cases of (3.12) have been used by Welch [33] and Kolodziejczyk [22].

$$\mathcal{D}(\mathbf{x}_{11}, \cdots, \mathbf{x}_{pn}) = \prod_{\nu} \mathcal{D}(\mathbf{x}_{1\nu}, \cdots, \mathbf{x}_{p\nu}).$$

Then if $n > n'$, the chance variables

$$\mathbf{a}_{(\gamma)}^{\prime i}, \quad \mathbf{w}_{ig}^{2}, \quad \mathbf{v}_i, \quad \mathbf{u}_i \tag{3.13}$$

are independently distributed, and have joint probability density

$$\prod_i \Big\{ G(u_i; n - n', \sigma^i) \cdot G(v_i; (m-1)(i-1), \sigma^i) \cdot \Big[\prod_g G(w_{ig}^2; 1, \sigma^i) \Big] \Big[\prod_\gamma G(a_{(\gamma)}^{\prime i}; n_\gamma - i + 1, \sigma^i) \Big] \Big\}. \tag{3.14}$$

If $n = n'$, the chance variables $\mathbf{a}_{(\gamma)}^{\prime i}$, $\mathbf{w}_{ig}^2$, and $\mathbf{v}_i$ are independently distributed and have a joint probability density which may be obtained from (3.14) *by suppressing the terms involving u_i and v_1.*

Proof. The chance variables (3.13) have been shown, in the corollaries to Theorems 10 and 11, to satisfy the conditions of Theorem 8. The proof is then completed by means of Theorem 9.*

The following corollaries are devoted to a consideration of the densities, moments and possible ranges of certain chance variables. Applications of the distributions are indicated.

COROLLARY 1. *Under the hypotheses of Theorem* 12, *it follows that*

$$\mathcal{D}(\mathbf{a}^i) = G(a^i; n - i + 1, \sigma^i),$$

and

$$\mathcal{D}(\mathbf{a}^1, \cdots, \mathbf{a}^p) = \prod_i \mathcal{D}(\mathbf{a}^i).$$

Furthermore $\mathcal{E}\{\prod_i (\mathbf{a}^i)^{\lambda_i/2}\} = \prod_i \bar{G}(\lambda_i, n-i+1, \sigma^i)$ *and* $0 < a^i < \infty$.

Proof. See Theorem 12 and Theorem 10, Corollary 2.

Since $a_j = a^1 \cdots a^j$, there may be deduced from Corollary 1 the distribution and moments of the generalized variance,† the ratio of the generalized variance to any of its principal minors,‡ and the quotient of any two of its principal minors. There may also be deduced the joint distribution and moments of the generalized variance and some of its principal minors.

* This distribution is a generalization of that obtained by Bartlett [1], p. 268, starting from Wishart's distribution. It may be remarked that the use of Theorem 6 in conjunction with Theorem 12 yields a similar generalization of the results of Bartlett's §2.

† See Wilks [34], pp. 477, 481.

‡ In the general case one has to transform $\mathcal{D}(\mathbf{a}^1, \cdots, \mathbf{a}^p)$ to find the probability that a ratio of products of the variables $\mathbf{a}^i$ exceeds a given quantity.

In the theory of relations between two sets of variables,* there occurs a statistic which measures the portion of the generalized variance of the first set of variables which is due to the variables of the second set.

Let $\zeta_{i\nu}=x_{i\nu}-z_{i\nu}$, and assume $\xi_{i\nu}$ to be the same function of the a_{ij} and $x_{i\nu}$ that $\zeta_{i\nu}$ is of the σ_{ij} and $x_{i\nu}$. Furthermore, let

$$
\begin{aligned}
g_{ij} &= \sum_{\nu} (x_{i\nu} - \zeta_{i\nu})(x_{j\nu} - \zeta_{j\nu}), \\
g_{ij1} &= \sum_{\nu} (x_{i\nu} - \xi_{i\nu})(x_{j\nu} - \xi_{j\nu}), \\
g_{ij2} &= \sum_{\nu} (\xi_{i\nu} - \zeta_{i\nu})(\xi_{j\nu} - \zeta_{j\nu}),
\end{aligned} \tag{3.15}
$$

and let

$$
\begin{aligned}
a_{ij1} &= g_{ij1}, \\
a_{ij2} &= \sum_{\nu} \xi_{i\nu}\xi_{j\nu}.
\end{aligned} \tag{3.16}
$$

It is well known that $g_{ij}=g_{ij1}+g_{ij2}$ and $a_{ij}=a_{ij1}+a_{ij2}$. Obviously

$$g_{ij} = a_{ij}, \qquad i, j = s+1, \cdots, p.$$

We shall define the residual variance in terms of what may be called the generalized covariances. Set $s \leqq p-s$, and let†

$$
W'_{js} = (-1)^{s-j+1} \begin{vmatrix}
0 & \cdots & 0 & g_{j,s+1} & \cdots & g_{jp} \\
\cdot & \cdots & \cdot & \cdot & \cdots & \cdot \\
0 & \cdots & 0 & g_{s,s+1} & \cdots & g_{sp} \\
g_{s+1,j} & \cdots & g_{s+1,s} & g_{s+1,s+1} & \cdots & g_{s+1,p} \\
\cdot & \cdots & \cdot & \cdot & \cdots & \cdot \\
g_{pj} & \cdots & g_{ps} & g_{p,s+1} & \cdots & g_{pp}
\end{vmatrix}, \qquad j = 1, \cdots, s.
$$

Then we note that from Theorem 11 and Theorem 8, the statistic

$$
B'_{js} = \frac{W'_{js}}{d} = \begin{vmatrix}
g_{jj2} & \cdots & g_{js2} \\
\cdot & \cdots & \cdot \\
g_{sj2} & \cdots & g_{ss2}
\end{vmatrix}
$$

is the generalized variance of $p-s$ observations of a vector variate which has $s-j+1$ components and hence has the distribution and moments of such a

* See Hotelling [20]. A knowledge of §§1–4 of that paper is assumed.

† By W_{js} we shall mean the same function of the a_{ij} that W'_{js} is of the g_{ij}. It is noted that W_{1s} is the numerator of the statistic q^2 defined by Hotelling [20], p. 333.

variance.* From Theorems 11 and 12 it follows that the parametric function estimated by B'_{1s} is σ/τ.

Assume $k>j$. Since

$$\frac{W'_{js}}{W'_{ks}} = \frac{B'_{js}}{B'_{ks}},$$

its distribution may readily be found in accordance with the preceding discussion of the distribution of the ratio of a generalized variance to one of its principal minors.

Since $\boldsymbol{B}'_{js}$ and $\boldsymbol{d}$ are independent chance variables, the chance variable $\boldsymbol{W}'_{js}$ is a product of independent variances and has the distribution and moments of such a product.†

COROLLARY 2. (a) *If $n>n'$, the chance variables $\boldsymbol{a}^j_{(\gamma)}$, $\boldsymbol{a}^j$ have density*

$$\prod_j G(a^j - a^j_{(1)} - \cdots - a^j_{(m)};\ n - n' + (j-1)(m-1),\ \sigma^j) \cdot \left[\prod_\gamma G(a^j_{(\gamma)};\ n_\gamma - j + 1,\ \sigma^j)\right]. \tag{3.17}$$

Furthermore

$$\mathcal{E}\left\{\prod_j (\boldsymbol{a}^j)^{\lambda_j/2}\left[\prod_\gamma (\boldsymbol{a}^j_{(\gamma)})^{\lambda_{j\gamma}/2}\right]\right\} = \prod_j \left\{\bar{G}\left(\lambda_j;\ n - j + 1 + \sum_\gamma \lambda_{j\gamma},\ \sigma^j\right)\cdot\left[\prod_\gamma \bar{G}(\lambda_{j\gamma};\ n_\gamma - j + 1,\ \sigma^j)\right]\right\}, \tag{3.18}$$

and

$$0 < a^j_{(\gamma)} < a^j - a^j_{(m)} - \cdots - a^j_{(\gamma+1)}, \qquad a^j_{(m+1)} = 0. \tag{3.19}$$

(b) *If $n=n'$, the chance variables*

$$\begin{aligned} &\boldsymbol{a}^1_{(\gamma)}, && \gamma = 1, \cdots, m-1,\\ &\boldsymbol{a}^j_{(\gamma)}, && j = 2, \cdots, p;\ \gamma = 1, \cdots, m,\\ &\boldsymbol{a}^j, && j = 1, \cdots, p, \end{aligned}$$

* The residual variance is defined to be $B_{1s}=W_{1s}/d$, and has the distribution of B_{1s} if and only if $g_{i,j+2}=a_{ij2}$, $(i,j=1,\cdots,s)$.

† If $s=p-s=2$, the statistic W_{12} is the square of the covariance tetrad difference. Hence if the two sets are independent, the distribution and moments of the covariance tetrad difference have been derived. It is noted that the variables of either set may be intercorrelated. The correlation tetrad difference will be discussed after Corollary 3. This analysis generalizes the discussions of Wilks [35], and Wishart [40].

have a joint probability density, joint moments, and possible ranges which are of the same form as (3.17), (3.18), *and* (3.19) *with the obvious differences caused by the omission of* $a'_{(m)}$.

Proof. The corollary is an immediate result of the theorem on the transformation of multiple integrals.

If $n > n'$, the joint density and moments of the chance variables $\boldsymbol{a}_{j(\gamma)}$, $\boldsymbol{a}_j$, $(j = 1, \cdots, i)$, may readily be obtained by transforming (3.17). The jacobian of the transformation

$$a_{j(\gamma)} = a^1_{(\gamma)} \cdots a^j_{(\gamma)}$$
$$a_j = a^1 \cdots a^j$$

is

$$\prod_g a_g^{-1} \left[\prod_\gamma a_{g(\gamma)}^{-1} \right],$$

and the possible ranges of the chance variables are

$$0 < a_{k(\gamma)} < \prod_{j=1}^{k} \left(\frac{a_j}{a_{j-1}} - \frac{a_{j(m)}}{a_{j-1(m)}} - \cdots - \frac{a_{j(\gamma+1)}}{a_{j-1(\gamma+1)}} \right),$$
$$0 < a_{k(\gamma)} < \left(a_k^{1/k} - a_{k(m)}^{1/k} - \cdots - a_{k(\gamma+1)}^{1/k} \right)^k. \tag{3.20}$$

It is noted that if $n = n'$, one may find the joint density of the chance variables

$$\boldsymbol{a}_{1(\gamma)}, \quad \boldsymbol{a}_1, \qquad \gamma = 1, \cdots, m-1,$$
$$\boldsymbol{a}_{j(\gamma)}, \quad \boldsymbol{a}_j, \quad j = 2, \cdots, i; \gamma = 1, \cdots, m,$$

in a similar fashion. The moments of the chance variables are most easily obtained by means of Lemma 1 and (3.18).

The inequalities (3.20) are consequences of the fact that

$$0 < a_{k(\gamma)} < \prod_{j=1}^{k} (a^j - a^j_{(m)} - \cdots - a^j_{(\gamma+1)})$$

and of the application of the procedure of finding the extrema of a function subject to several conditions.

From these distributions may be obtained the joint distribution of

$$\boldsymbol{a}_{i(1)}, \cdots, \boldsymbol{a}_{i(m)}, \boldsymbol{a}_i$$

by integrating the other variables over their possible ranges as stated in (3.20).

Let $c^j_{(\gamma)} = a^j_{(\gamma)}/a^j$. Then we may state the following corollary:

COROLLARY 3. (a). *If $n>n'$, then*

$$\begin{aligned}(3.21)\qquad &\mathcal{D}(\overset{i}{\mathbf{c}}_{(1)}, \cdots, \overset{i}{\mathbf{c}}_{(m)}) \\ &= D(\overset{i}{c}_{(1)}, \cdots, \overset{i}{c}_{(m)}; n_1-i+1, \cdots, n_m-i+1, n-n'+(i-1)(m-1)), \\ &\mathcal{D}(\overset{i}{\mathbf{c}}_{(1)}, \cdots, \overset{i}{\mathbf{c}}_{(m)}, \overset{i}{\mathbf{a}}) = \mathcal{D}(\overset{i}{\mathbf{a}}) \cdot \mathcal{D}(\overset{i}{\mathbf{c}}_{(1)}, \cdots, \overset{i}{\mathbf{c}}_{(m)}),\end{aligned}$$

and

$$\mathcal{D}(\overset{1}{\mathbf{c}}_{(1)}, \cdots, \overset{p}{\mathbf{c}}_{(m)}, \overset{1}{\mathbf{a}}, \cdots, \overset{p}{\mathbf{a}}) = \prod_i \mathcal{D}(\overset{i}{\mathbf{c}}_{(1)}, \cdots, \overset{i}{\mathbf{c}}_{(m)}, \overset{i}{\mathbf{a}}).$$

Furthermore

$$\begin{aligned}&\mathcal{E}\left\{ \prod_i \prod_\gamma (\overset{i}{\mathbf{c}}_{(\gamma)})^{\lambda_{i\gamma}/2} \right\} \\ &\quad = \prod_i \bar{D}(\lambda_{i1}, \cdots, \lambda_{im}; n_1-i+1, \cdots, n_m-i+1, n-n'+(m-1)(i-1)),\end{aligned}$$

and

$$0 < \overset{j}{c}_{(\gamma)} < 1 - \overset{j}{c}_{(m)} - \cdots - \overset{j}{c}_{(\gamma+1)}.$$

(b) *If $n=n'$, the chance variables*

$$\begin{aligned}&\overset{1}{\mathbf{c}}_{(\gamma)}, && \gamma = 1, \cdots, m-1, \\ &\overset{j}{\mathbf{c}}_{(\gamma)}, && j = 2, \cdots, p; \gamma = 1, \cdots, m,\end{aligned}$$

have a joint probability density, joint moments, and ranges which may be determined as above.

We shall call any statistic of the form

$$c_{i(\gamma)} = \frac{a_{i(\gamma)}}{a_i}$$

a generalized correlation ratio. Then, since

$$c_{j(\gamma)} = \overset{1}{c}_{(\gamma)} \cdots \overset{j}{c}_{(\gamma)},$$

we may obtain the distribution and moments of the generalized correlation ratio* and of the generalized Neyman-Pearson criteria.†

Applications of Corollary 3 occur in the theories of generalized periodo-

* See Wilks [34], p. 482.

† See Pearson and Wilks [29], and Wilks [34], p. 488. These criteria turn out to be functions of products of generalized correlation ratios.

gram analysis and relations between sets of variables. Of these, we first discuss generalized periodogram analysis.*

Let

$$z_{i\nu} = x_{i\nu} - \sum_{j=0}^{k} \alpha_{ij}\phi_j(t_\nu),$$

and let the functions $\phi_0(t), \cdots, \phi_k(t)$ be such that

$$\sum_{\nu} \phi_i(t_\nu)\phi_j(t_\nu) = \delta_{ij}, \qquad i, j = 0, \cdots, k.$$

It is well known that if

$$\mathcal{D}(z_{1\nu}, \cdots, z_{p\nu}) = N((z_\nu); (\sigma)), \quad \mathcal{D}(z_{11}, \cdots, z_{pn}) = \prod_{\nu} \mathcal{D}(z_{1\nu}, \cdots, z_{p\nu}),$$

and if

$$y_{i\mu} = \sum_{\nu} \phi_\mu(t_\nu)x_{i\nu}, \qquad i = 1, \cdots, p; \mu = 0, \cdots, k,$$

then

$$\mathcal{D}(y_{1\mu} - \alpha_{1\mu}, \cdots, y_{p\mu} - \alpha_{p\mu}) = N((y_\mu - \alpha_\mu); (\sigma)), \qquad \mu = 0, \cdots, k,$$

and

$$\mathcal{D}(y_{10} - \alpha_{10}, \cdots, y_{pk} - \alpha_{pk}) = \prod_{\mu=0}^{k} \mathcal{D}(y_{1\mu} - \alpha_{1\mu}, \cdots, y_{p\mu} - \alpha_{p\mu}).$$

Furthermore there exist linear functions

$$y_{i\mu} = \sum_{\nu} \phi_\mu(t_\nu)x_{i\mu}, \qquad \mu = k+1, \cdots, n-1,$$

such that

$$\mathcal{D}(y_{1\mu}, \cdots, y_{p\mu}) = N((y_\mu); (\sigma)),$$

$$\mathcal{D}(y_{1,k+1}, \cdots, y_{p,n-1}) = \sum_{\mu=k+1}^{n-1} \mathcal{D}(y_{1\mu}, \cdots, y_{p\mu}),$$

and†

$$\mathcal{D}(y_{10} - \alpha_{10}, \cdots, y_{pn}) = \mathcal{D}(y_{10} - \alpha_{10}, \cdots, y_{pk} - \alpha_{pk}) \cdot \mathcal{D}(y_{1,k+1}, \cdots, y_{p,n-1}).$$

We consider our sample to contain an odd number‡ n, $(n = 2n' + 1)$, of observations, and let

---

* See Fisher [14], and Greenstein [17], for discussions of the Schuster periodogram in one variable.

† This follows from Corollary 2 of Theorem 7.

‡ The adjustment to be made if n' is even will be obvious. See Fisher [14], p. 60.

$$\phi_0(t_\nu) = n^{-1/2},$$

$$\phi_{2j}(t_\nu) = \left(\frac{2}{n}\right)^{1/2} \sin \frac{2\pi\nu j}{n},$$

$$\phi_{2j-1}(t_\nu) = \left(\frac{2}{n}\right)^{1/2} \cos \frac{2\pi\nu j}{n}, \qquad j = 1, \cdots, n'; \nu = 1, \cdots, n.$$

If $k < n-1$, it may be shown that a Neyman-Pearson λ criterion* for the hypothesis

$$\alpha_{ij} = 0, \qquad i = 1, \cdots, p; j = 1, \cdots, k,$$

is given by a generalized correlation ratio

$$\lambda = \frac{\left| \sum_{\nu=k+1}^{n-1} y_{i\nu} y_{j\nu} \right|}{\left| \sum_{\nu=1}^{n-1} y_{i\nu} y_{j\nu} \right|}, \qquad i, j = 1, \cdots, p.$$

However, if $k = n-1$, or if it is desired that the hypothesis be rejected if and only if at least one of the statistics

$$\bar{c}_{p(2\gamma)} = \frac{\left| \sum_{\nu=1}^{n-1} y_{i\nu} y_{j\nu} - y_{i,2\gamma-1} y_{j,2\gamma-1} - y_{i,2\gamma} y_{j,2\gamma} \right|}{\left| \sum_{\nu=1}^{n-1} y_{i\nu} y_{j\nu} \right|}, \tag{3.22}$$

$$\gamma = 1, \cdots, n'; i, j = 1, \cdots, p,$$

differs significantly from one, some other approach to the problem of testing the hypothesis that no period exists must be given.

Inasmuch as the joint distributions of statistics such as $\bar{c}_{p(2\gamma)}$ is quite complicated,† we shall extend Fisher's distribution‡ only to the case $p = 2$, and shall use, in doing this, not the statistics (3.22) but the statistics

$$c_{2(2\gamma)} = \frac{\left| y_{i,2\gamma-1} y_{j,2\gamma-1} + y_{i,2\gamma} y_{j,2\gamma} \right|}{\left| \sum_{\nu=1}^{n-1} y_{i\nu} y_{j\nu} \right|}, \qquad \gamma = 1, \cdots, n'; i, j = 1, 2. \tag{3.23}$$

* See Neyman and Pearson [27], and Wilks [34], p. 485.

† They may be found by means of the distributions of Corollary 3 and variables having a normal multivariate distribution. Some notion of the complications may be gained by examining the distribution for $p = 1$.

‡ R. A. Fisher [14], p. 57. The solution of a similar problem, if the variance and covariance parameters are known, is carried through in the same way.

The joint probability density of a subset $c_{2(2\gamma_1)}, c_{2(2\gamma_2)}, \cdots, c_{2(2\gamma_k)}$ of the set (3.23), is, from (3.21),

$$\frac{\Gamma(n')\Gamma(n'-1/2)}{\pi^{k/2}\Gamma(n'-k)\Gamma(n'-(k+1)/2)}\left\{\prod_{\delta=1}^{k} c_{2(2\gamma_\delta)}^{-1/2}\right\}$$
$$\cdot\left\{\int_{c_{2(2\gamma_k)}}^{1}\cdots\int_{c_{2(2\gamma_1)}}^{1-c_{1(2\gamma_k)}-\cdots-c_{1(2\gamma_1)}}\left[\prod_{\delta=1}^{k} c_{1(2\gamma_\delta)}^{-1/2}\right]\right.$$
$$\cdot(1-c_{1(2\gamma_1)}-\cdots-c_{1(2\gamma_k)})^{n'-k+1}$$
$$\left.\cdot\left(1-\frac{c_{2(2\gamma_1)}}{c_{1(2\gamma_1)}}-\cdots-\frac{c_{2(2\gamma_k)}}{c_{1(2\gamma_k)}}\right)^{n'-(k+1)/2-1} dc_{1(2\gamma_1)}\cdots dc_{1(2\gamma_k)}\right\}$$

while the joint density of the entire set is

$$\frac{\Gamma(n')\Gamma(n'-1/2)}{\pi^{k/2}\Gamma((k-1)/2)}\left\{\prod_{\gamma=1}^{n'} c_{2(2\gamma)}^{-1/2}\right\}\left\{\int_{c_{2(2n'-2)}}^{1}\cdots\int_{c_{2(2)}}^{1-c_{1(2n'-2)}-\cdots-c_{1(4)}}\left(\prod_{\gamma=1}^{n'-1} c_{1(2\gamma)}^{-1/2}\right)\right.$$
$$\cdot(1-c_{1(1)}-\cdots-c_{1(2n'-2)})^{-1/2}\left(1-\frac{c_{2(2n')}}{1-c_{1(2)}-\cdots-c_{1(2n'-2)}}\right.$$
$$\left.\left.-\frac{c_{2(2n'-2)}}{c_{1(2n'-2)}}-\cdots-\frac{c_{2(2)}}{c_{1(2)}}\right)^{(n'-1)/2-1} dc_{1(2)}\cdots dc_{1(2n'-2)}\right\}.$$

If we denote the probability that

$$c_{2(2\gamma_i)} > \theta_{\gamma_i}, \qquad i = 1, \cdots, k,$$

for each selection of k integers from $1, 2, \cdots, n'$, $(k=1, 2, \cdots, n')$ by $P(\theta_{\gamma_1}, \cdots, \theta_{\gamma_k})$, then the probability that at least j of the events

$$c_{2(2\gamma)} > \theta_\gamma, \qquad \gamma = 1, \cdots, n',$$

occur is*

$$P^{(j)} = \sum_{\nu=0}^{n'-j} (-1)^\nu c_\nu^{j+\nu-1} S_{j+\nu} \tag{3.24}$$

where

$$S_{j+\nu} = \sum P(\theta_{\gamma_1}, \cdots, \theta_{\gamma_{j+\nu}})$$

and the summation is for all selections of $j+\nu$ integers from $1, \cdots, n$.†

From (3.24), we can find the probability that the maximum of the statistics $c_{2(2\gamma)}$ exceeds a given quantity g, by letting $j=1$ and $\theta_1 = \cdots = \theta_{n'} = g$. It is, however, a question whether only the maximum need be considered or

* See, for example, [25], p. 162.

† If $\theta_1=\theta_2=\cdots=\theta_n$, then $S_{j+\nu}=C_{j+\nu}^{n'}P(\theta_1, \cdots, \theta_{j+\nu})$.

whether each statistic should first be tested and then, if h are significant, the probability that the h greatest of the statistics should exceed a given quantity g should be found.

We note that by using the corollaries to Theorem 13, other distributions which may be used to test hypotheses in generalized periodogram analysis may be obtained for $p>2$. However we shall not examine these here.

We now derive the joint distribution of q^2 and z* and the joint distribution of the two sample canonical correlations,† which exist if $s=2$, when the two sets are not completely independent.

The degree of independence which we admit is expressed by the assumption that there exist s linear functions $y_{1\nu}, \cdots, y_{s\nu}$ of $x_{1\nu}, \cdots, x_{s\nu}$, and $p-s$ linear functions $y_{s+1,\nu}, \cdots, y_{p\nu}$ of $x_{s+1,\nu}, \cdots, x_{p\nu}$, such that the probability density of the chance variables

$$y_{i\nu}, \qquad i = 1, \cdots, p; \nu = 1, \cdots, n, \tag{3.25}$$

is

$$(2\pi)^{pn/2}(1-\rho^2)^{n/2} \exp\left[-\frac{1}{2(1-\rho^2)}\sum_{\nu}\left\{\sum_{i} y_{i\nu}^2 - 2\rho y_{1\nu}y_{s+1,\nu}\right\}\right].$$

We first note that if $\rho=0$, the joint probability density of

$$q^2 = \frac{W_{1s}}{a_s d} = \frac{a_{s(2)}}{a_{(s)}},$$

and

$$z = \frac{a_p}{a_s d} = \frac{a_{s(1)}}{a_s},$$

where‡

$$a_{s(1)} = |a_{ij1}|,$$
$$a_{s(2)} = |a_{ij2}|, \qquad i, j = 1, \cdots, s,$$

may be obtained from Theorem 12, Corollary 3, since both q^2 and z are invariant under the transformation, which we have assumed, carrying the chance variables $x_{i\nu}$ into the chance variables (3.25), $(\rho=0)$.

It follows from Theorem 9 and Corollary 3, that if we denote a_{112}/a_{11} by R^2, then the chance variables

* Hotelling ([20], p. 333) defined q^2 and obtained on page 354 its distribution when one non-vanishing canonical correlation parameter exists. The distribution of z in that case may readily be obtained by the same procedure as that by which we shall obtain the joint distribution of q^2 and z.

† Hotelling ([20], p. 375) derived this distribution assuming $s=p-s=2$ and complete independence of sets.

‡ a_{ij1} and a_{ij2} are defined in (3.16).

$$R^2,\ c^i_{(1)},\ c^i_{(2)}, \qquad i = 1, \cdots, s,$$

have density

$$\left\{\prod_{j=1}^{s} \frac{\Gamma((n-j+1)/2)}{\Gamma((p-s-j+1)/2)\cdot\Gamma((n-p+s-j+1)/2)\cdot\Gamma((j-1)/2)}\right\}$$

$$\cdot(1-\rho^2)^{n/2}(R^2)^{(p-s)/2-1}(1-R^2)^{(n-p+s)/2-1}F(n/2, n/2, (p-s)/2, \rho^2R^2) \tag{3.26}$$

$$\cdot\left\{\prod_{j=2}^{s}(c^j_{(1)})^{(n-p+s-j+1)/2-1}(c^j_{(2)})^{(p-s-j+1)/2-1}(1-c^j_{(1)}-c^j_{(2)})^{(j-1)/2-1}\right\},$$

where $F(n/2, n/2, (p-s)/2, \rho^2R^2)$ is the hypergeometric function.* If $s=2$, then $q^2=R^2\cdot c^2_{(2)}$, $z=(1-R^2)c^2_{(1)}$, and

$$\mathcal{D}(q^2, z) = \frac{\Gamma(n)(1-\rho^2)^{n/2}}{32\pi\Gamma(p-2)\Gamma(n-p-2)}\, z^{(n-p+1)/2-1}(q^2)^{(p-3)/2-1}$$

$$\cdot\int_0^1 F(n/2, n/2, (p-2)/2, \rho^2[x(1-z-q^2)+q^2]) \tag{3.27}$$

$$\cdot\left[x(1-x)+\frac{q^2z}{1-z-q^2}\right]^{-1}dx.$$

Then the general distribution may be similarly obtained as a multiple integral. The joint moments of q^2 and z are obtainable directly from (3.26).

Another statistic which may occasionally be of interest is

$$D = \frac{q^2}{z}$$

the distribution of which may be obtained as above. The factors of D are ratios of independent variances rather than correlation ratios if the sets are independently distributed.

The joint distribution of the two sample canonical correlation coefficients may be evaluated by means of the relations†

$$q^2 = r_1^2 r_2^2$$
$$z = (1-r_1^2)(1-r_2^2).$$

Since $\partial(q^2, z)/\partial(r_1^2, r_2^2) = r_1^2 - r_2^2$, if $r_1^2 > r_2^2$, it follows that the chance variables r_1^2 and r_2^2 have a density which may be obtained by transforming (3.27).

---

* The details consist of letting $c=[a_{11}a_{s+1,s+1}-(a_{1,s+1})^2]/a_{s+1,s+1}$, noting that $a_{1,s+1}=a_{11}^{1/2}a_{s+1,s+1}^{1/2}\cdot(1-c/a_{11})^{1/2}$, using Corollary 3 and Theorem 9, and then employing the same procedure as that adopted by R. A. Fisher [13].

† Hotelling [19], p. 335.

The first measure of relations between two sets of variables, both of which contained more than one variable, was the tetrad difference. However, as will be shown, the probability density of the tetrad difference depends on the "internal" correlation parameters even when the two sets are assumed to be independent. It is this fact which makes it necessary that q^2 be used instead of the tetrad difference unless the correlation parameters are known.

To discuss the statistic

$$T_{1s} = \frac{W_{1s}}{a_{11}\ a_{22} \cdots a_{pp}}, \tag{3.28}$$

which is a generalization of the tetrad difference, we first note that if all the variables are independent, or if we consider the statistic

$$T'_{1s} = \frac{W'_{1s}}{g_{11}\ g_{22} \cdots g_{pp}}$$

in either of these two cases, the statistic is a product of independent determinants of correlation coefficients and has the distribution and moments of such a product.*

We now assume that $s = p - s = 2$ and derive the distribution of the square of the correlation tetrad difference under the assumption that the two variables of each set are correlated but that the two sets are independent.

Let the chance variables $\mathbf{x}_{iv}$, $(i = 1, 2, 3, 4; v = 1, \cdots, n)$, have probability density†

$$(2\pi)^{-2n}(1 - \overset{2}{\rho}_{12})^{-n/2}(1 - \overset{2}{\rho}_{34})^{-n/2} \cdot \exp\left[- \frac{1}{2(1 - \rho_{12}^2)} \left\{ \sum_v (\overset{2}{x}_{1v} - 2\rho_{12}x_{1v}x_{2v} + \overset{2}{x}_{2v}) \right\} - \frac{1}{2(1 - \rho_{34}^2)} \left\{ \sum_v (\overset{2}{x}_{3v} - 2\rho_{34}x_{3v}x_{4v} + \overset{2}{x}_{4v}) \right\} \right].$$

Now by (3.28),

$$T_{12} = \frac{a_{112}}{a_{11}} \cdot \frac{\overset{2}{a}_{(2)}}{a_{22}} \cdot (1 - \overset{2}{r}_{34}).$$

* If all the variables are independent, then the p-rowed determinant of correlation coefficients is the product of $p-1$ independent multiple correlation coefficients and has the corresponding distribution and moments. Wilks, [34], p. 492, first derived this distribution by the moment method. Certain more general distributions, which do not depend on the vanishing of all the correlation parameters may easily be derived.

† The assumption $\sigma_{ii} = 1$, $(i = 1, \cdots, 4)$, involves no loss of generality since T_{1s} is invariant under the transformation $y_{iv}\sigma_{ii} = x_{iv}$.

Let

$$R^2 = \frac{a_{112}}{a_{11}}, \qquad V_1 = \frac{a^2_{(2)}}{a_{22}}, \qquad V_2 = (1 - r^2_{34}).$$

Then R^2, V_1, and V_2 are independent, and

$$\begin{aligned}\mathcal{D}(R^2, V_1, V_2) = {} & \frac{2[\Gamma(n/2)]^3(1-\rho^2_{12})^{n/2}(1-\rho^2_{34})^{n/2}}{\pi\Gamma((n-3)/2)\Gamma((n-2)/2)\Gamma((n-1)/2)} \cdot (1-V_1)^{(n-4)/2} \\ & \cdot (R^2)^{1/2}(1-R^2)^{(n-5)/2} F(n/2, n/2, 3/2, \rho^2_{12}R^2) \\ & \cdot V_2{}^{(n-3)/2}(1-V_2)^{1/2} F(n/2, n/2, 1/2, \rho^2_{34}V_2).\end{aligned}$$

From this may be found at once the moments and distribution of T_{12}. The distribution and moments of T_{1s} may also be found.

Let $b_{ij\gamma} = \sum_{t=1}^{\gamma} a_{ijt}$, and let $b_{i(\gamma)}$, $b^i_{(\gamma)}$ be defined as are $a_{i(\gamma)}$ and $a^i_{(\gamma)}$. Let $b'_{ij\gamma}$, and so on, be similar functions of the variables y_{kv}. Assume that $b'^i_0 = 0$, and let

$$u_{i(\gamma)} = b'^i_{(\gamma)} - b'^i_{(\gamma-1)}.$$

We shall assume that $n = n_1 + \cdots + n_m$. This involves no loss of generality, for if $n > n_1 + \cdots + n_m$, we may extend the system $b'_{ij\gamma}$ to include b'_{ijm+1}.

THEOREM 13. *Let* $\mathcal{D}(\mathbf{x}_{1v}, \cdots, \mathbf{x}_{pv}) = N((x_v); (\sigma))$, *and let* $\mathcal{D}(\mathbf{x}_{11}, \cdots, \mathbf{x}_{pn}) = \prod_v \mathcal{D}(\mathbf{x}_{1v}, \cdots, \mathbf{x}_{pv})$. *Then*

$$\begin{aligned}\mathcal{D}(\boldsymbol{u}_{i(1)}) &= G(u_{i(1)}; n_1 - i + 1, \sigma^i), \\ \mathcal{D}(\boldsymbol{u}_{i(\gamma)}) &= G(u_{i(\gamma)}; n_\gamma, \sigma^i), \qquad \gamma = 2, \cdots, m, \\ \mathcal{D}(\boldsymbol{w}^2_{ig}) &= G(w^2_{ig}; 1, \sigma),\end{aligned}$$

and

$$\mathcal{D}(\boldsymbol{u}, \boldsymbol{w}^2) = \prod_i \left[\prod_\gamma \mathcal{D}(\boldsymbol{u}_{i(\gamma)})\right]\left[\prod_g \mathcal{D}(\boldsymbol{w}^2_{ig})\right].$$

Proof. See Theorem 11.

COROLLARY 1.

$$(3.29) \qquad \mathcal{D}(\boldsymbol{b}^j_{(1)}, \cdots, \boldsymbol{b}^j_{(m)}) = G(b^j_{(1)}; n_1 - j + 1, \sigma^j) \prod_{\gamma=2}^{m} G(b^j_{(\gamma)} - b^j_{(\gamma-1)}; n_\gamma \sigma^j)$$

and

$$\mathcal{D}(\boldsymbol{b}^1_{(1)}, \cdots, \boldsymbol{b}^p_{(m)}) = \prod_j \mathcal{D}(\boldsymbol{b}^1_{(1)}, \cdots, \boldsymbol{b}^j_{(m)}).$$

Furthermore

$$\mathcal{E}\left\{\prod_j \prod_\gamma (\boldsymbol{b}_{(\gamma)}^j)^{\lambda_{j\nu}/2}\right\}$$

$$= \prod_j \prod_\gamma \overline{G}\left(\lambda_{j\gamma}; \sum_{\delta=1}^{\gamma} (n_\delta + \lambda_{j\delta}) - j + 1 - \lambda_{j\gamma}, \sigma^j\right) \tag{3.30}$$

and

$$0 < b_{(1)}^j < b_{(2)}^j < \cdots < b_{(m)}^j < \infty .$$

The proof is omitted.

Let $d_{(\gamma)}^j = b_{(\gamma)}^j / b_{(m)}^j$. The probability density and joint moments of the chance variables $\boldsymbol{d}_{(\gamma)}^j$ may be obtained by transforming (3.29) and (3.30). Furthermore

$$0 < d_{(1)}^j < \cdots < d_{(m-1)}^j < 1.$$

Let $\omega_{(\gamma)}^j = b_{(\gamma)}^j / b_{(\gamma+1)}^j$. We can then state the following corollary:

COROLLARY 2.

$$\mathcal{D}(\boldsymbol{\omega}_{(\gamma)}^j) = D(\omega_{(\gamma)}^j; n_1 + \cdots + n_\gamma - j + 1, n_{\gamma+1}),$$

$$\mathcal{D}(\boldsymbol{\omega}_{(1)}^1, \cdots, \boldsymbol{\omega}_{(m-1)}^p) = \prod_j \prod_{\gamma=1}^{m-1} \mathcal{D}(\boldsymbol{\omega}_{(\gamma)}^j),$$

and

$$\mathcal{D}(\boldsymbol{\omega}_{(1)}^1, \cdots, \boldsymbol{\omega}_{(m-1)}^p, \boldsymbol{a}^1, \cdots, \boldsymbol{a}^p) = \mathcal{D}(\boldsymbol{\omega}_{(1)}^1, \cdots, \boldsymbol{\omega}_{(m-1)}^p) \cdot \mathcal{D}(\boldsymbol{a}^1, \cdots, \boldsymbol{a}^p).$$

The proof is omitted. The moments and possible ranges of the chance variables $\omega_{(\gamma)}^j$ are too well known to require statement here.

We are now able to derive several additional distributions of the generalized analysis of variance.

To do so, we need only note that

$$b_{j(\gamma)} = \prod_{i=1}^{j} b_{(\gamma)}^i,$$

and that the probability density and moments of the statistics $\boldsymbol{b}_{j(\gamma)}$ may therefore be obtained from Theorem 13, Corollary 1.

The probability density and moments of the statistics $\boldsymbol{d}_{j(\gamma)}$, where

$$d_{j(\gamma)} = \frac{b_{j(\gamma)}}{b_{j(m)}},$$

may be similarly derived.

It may be remarked that the independence of the chance variables $\omega_{(\gamma)}^j$

permits the distributions depending on them to be expressed in terms of previously obtained distributions.

Distributions derived from Theorem 13 may be used in testing various hypotheses in generalized analysis of variance, comparative analysis,* and relations between sets of variables. The method to be used should be sufficiently clear to make any further transformations here unnecessary.

Bibliography

1. M. S. Bartlett, *On the theory of statistical regression*, Proceedings of the Royal Society of Edinburgh, vol. 53 (1933), pp. 260–283.

2. M. S. Bartlett, *The problem in statistics of testing several variances*, Proceedings of the Cambridge Philosophical Society, vol. 30 (1934), pp. 164–169.

3. M. S. Bartlett, *The vector representation of a sample*, Proceedings of the Cambridge Philosophical Society, vol. 30 (1934), pp. 327–340.

4. M. S. Bartlett and J. Wishart, *The generalized product moment distribution in a normal system*, Proceedings of the Cambridge Philosophical Society, vol. 29 (1933), pp. 260–270.

5. S. Bochner, *Vorlesungen über Fouriersche Integrale*, Akademische Verlagsgesellschaft, Leipzig, 1932.

6. W. G. Cochran, *The distribution of quadratic forms in a normal system, with applications to the analysis of covariance*, Proceedings of the Cambridge Philosophical Society, vol. 30 (1934), pp. 178–191.

7. H. Cramer, *Random Variables and Probability Distributions*, Cambridge Tracts in Mathematics and Mathematical Physics, no. 36, London, 1937.

8. G. Darmois, *Sur les lois de probabilité à estimation exhaustive*, Comptes Rendus de l'Académie des Sciences, Paris, vol. 200 (1935), pp. 1265–1266.

9. R. A. Fisher, *On the mathematical foundations of theoretical statistics*, Philosophical Transactions of the Royal Society of London, (A), vol. 222 (1921), pp. 309–368.

10. R. A. Fisher, *The influence of rainfall on the yield of wheat at Rothamstead*, Philosophical Transactions of the Royal Society of London, (B), vol. 213 (1924), pp. 89–142.

11. R. A. Fisher, *On a distribution yielding the error functions of several well known statistics*, Proceedings of the International Mathematical Congress, Toronto, 1924, pp. 805–812.

12. R. A. Fisher, *Applications of "Student's" distribution*, Metron, vol. 5 (1926), pp. 90–104.

13. R. A. Fisher, *The general sampling distribution of the multiple correlation coefficient*, Proceedings of the Royal Society of London, (A), vol. 121 (1928), pp. 654–673.

14. R. A. Fisher, *Tests of significance in harmonic analysis*, Proceedings of the Royal Society of London, (A), vol. 125 (1929), pp. 54–60.

15. R. A. Fisher, *Two new properties of mathematical likelihood*, Proceedings of the Royal Society of London, (A), vol. 144 (1934), pp. 285–307.

16. R. A. Fisher, *Statistical Methods for Research Workers*, 6th edition. Oliver and Boyd, London, 1936.

17. B. Greenstein, *Periodogram analysis with special application to business failures in the United States, 1867–1932*, Econometrica, vol. 3 (1935), pp. 170–198.

18. H. Hotelling, *The generalization of Student's ratio*, Annals of Mathematical Statistics, vol. 2 (1931), pp. 170–198.

19. H. Hotelling, *Analysis of a complex of statistical variables into principal components*, Journal of Educational Psychology, vol. 24 (1933), pp. 417–441, 498–520.

* We use comparative analysis to denote the type of statistical analysis of which periodogram analysis is an example, that is, statistical analysis in which the effects of certain types of selection must be excluded.

20. H. Hotelling, *Relations between two sets of variables*, Biometrika, vol. 28 (1936), pp. 321–377.

21. A. Kolmogoroff, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Ergebnisse der Mathematik, vol. 2, no. 3.

22. S. Kolodziejczyk, *On an important class of statistical hypotheses*, Biometrika, vol. 27 (1935), pp. 161–190.

23. B. O. Koopman, *On distributions admitting a sufficient statistic*, these Transactions, vol. 39 (1936), pp. 399–410.

24. S. Kullback, *An application of characteristic functions to the distribution problem of statistics*, Annals of Mathematical Statistics, vol. 5 (1934), pp. 264–307.

25. W. G. Madow, *Contributions to the theory of comparative statistical analysis. I Fundamental theorems of comparative analysis*, Annals of Mathematical Statistics, vol. 8 (1937), pp. 159–176.

26. J. Neyman, *Su un teorema concernente le cosidette statistische sufficiente*, Giornale dell' Istituto Italiano degli Attuari, vol. 6 (1935), pp. 320–334.

27. J. Neyman and E. S. Pearson, *On the use and interpretation of certain test criteria for purposes of statistical inference*, Biometrika, vol. 20^A (1928), pp. 175–240, 264–294.

28. J. Neyman and E. S. Pearson, *Sufficient Statistics and Uniformly Most Powerful Tests*, Statistical Research Memoirs, Cambridge University Press, London, vol. 1, 1936.

29. E. S. Pearson and S. S. Wilks, *Methods of statistical analysis appropriate for k samples of two variables*, Biometrika, vol. 25 (1933), pp. 363–377.

30. E. S. Pearson and J. Neyman, see [27] and [28].

31. S. Saks, *Theory of the Integral*, 2d edition, Stechert, New York, 1937.

32. H. W. Turnbull, *The Theory of Determinants, Matrices, and Invariants*, Blackie and Son, London, 1926.

33. B. L. Welch, *Some problems in the analysis of regression among k samples of two variables*, Biometrika, vol. 27 (1935), pp. 145–160.

34. S. S. Wilks, *Certain generalizations in the analysis of variance*, Biometrika, vol. 24 (1932), pp. 472–494.

35. S. S. Wilks, *The standard error of a tetrad in samples from a normal population of independent variables*, Proceedings of the National Academy of Sciences, vol. 18 (1932), pp. 562–565.

36. S. S. Wilks and E. S. Pearson, see [29].

37. S. S. Wilks, *Moment-generating operators for determinants of product moments in samples from a normal system*, Annals of Mathematics, (2), vol. 35 (1933), pp. 312–340.

38. S. S. Wilks, *On the independence of k sets of normally distributed statistical variables*, Econometrika, vol. 3 (1935), pp. 309–326.

39. J. Wishart, *The generalized product moment distribution in samples from a normal multivariate population*, Biometrika, vol. 20^A (1928), pp. 32–52.

40. J. Wishart, *Sampling errors in the theory of two factors*, British Journal of Psychology, vol. 19 (1928), pp. 32–52.

41. J. Wishart and M. S. Bartlett, see [4].

COLUMBIA UNIVERSITY,
NEW YORK, N. Y.

A CONTINUOUS FUNCTION WITH NO UNILATERAL DERIVATIVES*

BY
ANTHONY P. MORSE

1. Introduction

A. S. Besicovitch† has given a construction of an even continuous function B for which he asserts the property $D_+B(x) < D^+B(x)$, $(-1 < x < 1)$, and, in consequence of the evenness of B, the companion property $D_-B(x) < D^-B(x)$, $(-1 < x < 1)$; that is, B has nowhere a unilateral derivative finite or infinite. Later E. D. Pepper‡ examined this same function. Some readers have found difficulty in following the reasoning employed by both these authors, and it may be that in some minds doubt as to the existence of such a function has been raised by the theorem of S. Saks§ to the effect that the "functions of Besicovitch" constitute a set of only *first* category in the space C of continuous functions.‖

In the present paper we, like Besicovitch, associate with a function having dense intervals of constancy another such function. The method of association used here, however, is purely arithmetic and differs essentially from that of Besicovitch. The function F which it is our purpose to define and investigate is even and continuous on the open interval $(-1, 1)$ and has the property

$$\liminf_{\xi \to x+} \left| \frac{F(\xi) - F(x)}{\xi - x} \right| < \limsup_{\xi \to x+} \left| \frac{F(\xi) - F(x)}{\xi - x} \right| = \infty, \qquad -1 < x < 1,$$

which, incidentally, the function B does not possess.

We remark that it is our intention to order the subsequent material so as to facilitate the reading of the proof of Theorem 5.2. Should the reader wish to acquaint himself at once with the formal definition of F, he may do so by examining §2 and Definitions 3.1, 3.3, 4.2, and 5.1.

* Presented to the Society, December 30, 1937; received by the editors December 18, 1937. The author wishes to express his gratitude to E. M. Beesley of Brown University for assistance in preparation of the manuscript and for suggestions concerning presentation of the material.

† Besicovitch, *Discussion der stetigen Funktionen im Zusammenhang mit der Frage über ihre Differentiierbarkeit*, Bulletin de l'Académie des Sciences de Russie, vol. 19 (1925), pp. 527–540.

‡ Pepper, *On continuous functions without a derivative*, Fundamenta Mathematicae, vol. 12 (1928), pp. 244–253.

§ Saks, *On the functions of Besicovitch in the space of continuous functions*, Fundamenta Mathematicae, vol. 19 (1932), pp. 211–219.

‖ It may be recalled that Banach, *Über die Baire'sche Kategorie gewisser Funktionenmengen*, Studia Mathematica, vol. 3 (1931), p. 174, has shown that in C the set of functions having at no point a *finite* unilateral derivative is of *second* category.

2. Notation and conventions

We employ the symbol $[x_1, x_2]$ to denote a closed interval, the symbol (x_1, x_2) to denote an open interval. The latter notation will be understood to imply that $x_1 < x_2$. By the word *set* we understand a set of real numbers, by the word *function* a function on a set to a set. The letters n, N, ν are permitted to assume integer values only. We denote the outer Lebesgue measure of a set A by $|A|$.

We shall find it convenient to refer to an open interval (α, β) as an *interval of* a set A if $(\alpha, \beta) \subset A$ with α and β elements of the closure of the complement of A. In case A is open this means α and β are not elements of A.

In connection with a function f we employ the notations:

$$K(f) = \text{the interior of the set } \underset{x}{E}\left[\lim_{h\to 0}[f(x+h)-f(x)]/h = 0\right],$$

$$H(f) = \text{the complement of } K(f) \text{ with respect to the domain of } f,$$

$$Z(f) = \underset{x}{E}[f(x) = 0], \qquad P(f) = \underset{x}{E}[f(x) > 0].$$

We also agree that if A is a set and $r \geqq 0$, then the set $A^{(r)}$ shall be defined as follows: $x \in A^{(r)}$ if $x = \pm 1$ or if there is an interval (a, b) such that $x \in (a, b) \subset A$ with $b - a > r$.

3. The sequence $\{\lambda_n\}$ and the function θ

3.1. Definition. *$\{\lambda_n\}$ is the sequence, defined on all integers n, for which*

$$\lambda_n = 1/2 + n/2(|n| + 3).$$

In addition to the fact that $\lambda_n \to 1$ or 0 according as $n \to +\infty$ or $-\infty$, the properties of $\{\lambda_n\}$ for which we shall have use are embodied in the following lemma:

3.2. Lemma. *For each integer n, $\lambda_n + \lambda_{-n} = 1$ and $0 < \lambda_{n+1} - \lambda_n < \lambda_n^2 < \lambda_n$.*

Proof. The first relation being evident, we prove the second. For each integer n, $|n|(n+1) - |n+1|n = 0$, and we have

$$\begin{aligned}\lambda_{n+1} - \lambda_n &= \frac{1}{2}\left(\frac{n+1}{|n+1|+3} - \frac{n}{|n|+3}\right)\\ &= \frac{1}{2}\left(\frac{|n|(n+1) + 3(n+1) - n|n+1| - 3n}{(|n+1|+3)(|n|+3)}\right)\\ &= \frac{3}{2(|n+1|+3)(|n|+3)} = \left(\frac{3}{2(|n|+3)}\right)^2 \cdot \frac{2(|n|+3)}{3(|n+1|+3)}\end{aligned}$$

$$\leqq \left(\frac{3+n+|n|}{2(|n|+3)}\right)^2 \cdot \frac{2(|n+1|+3+1)}{3(|n+1|+3)}$$

$$= (\lambda_n)^2 \cdot \frac{2}{3}\left(1+\frac{1}{|n+1|+3}\right)$$

$$\leqq (\lambda_n)^2(2/3)(1+1/3) = (\lambda_n)^2(8/9) < (\lambda_n)^2 < \lambda_n.$$

Since λ_n converges to 1 or 0 according as $n\to+\infty$ or $-\infty$, and since $0<\lambda_n<\lambda_{n+1}<1$ for each integer n (see 3.2), it follows that there is a unique function θ determined by the following definition:

3.3. DEFINITION. *With $\gamma_n=\lambda_n$ for n odd and $\gamma_n=(\lambda_n)^{1/2}$ for n even and $\mathfrak{H}$ a particular non-dense closed set which enjoys the property**

$$|\,[\lambda_n - h, \lambda_n]\cdot\mathfrak{H}\,| > 0, \quad |\,[\lambda_n, \lambda_n + h]\cdot\mathfrak{H}\,| > 0, \quad h > 0, -\infty < n < \infty,$$

θ is the function on $(0, 2)$ defined by the following relations:

$$\theta(x) = \gamma_n + (\gamma_{n+1} - \gamma_n)\,|\,[\lambda_n, x]\cdot\mathfrak{H}\,|\,/\,|\,[\lambda_n, \lambda_{n+1}]\cdot\mathfrak{H}\,|, \qquad \lambda_n \leqq x \leqq \lambda_{n+1},$$

$$\theta(1) = 1, \qquad \theta(x) = \theta(2-x), \qquad 0 < x < 2.$$

3.4. THEOREM. *The function θ has the following properties:*

(a) *θ is continuous on $(0, 2)$ and*

$$\lim_{\xi\to 0+} \theta(\xi) = 0 < x/2 \leqq \theta(x) \leqq (x+3)/4 \leqq \theta(1) = 1, \qquad 0 < x < 1;$$

(b) *$K(\theta)$† is dense in $(0, 2)$, while $\lambda_n \in H(\theta)$ for each integer n;*

(c) $\lim_{\xi\to x+} |\,[\theta(\xi) - \theta(x)]/(\xi - x)\,| < \infty, \qquad 0 < x < 2;$

(d) *if (α, β) is an interval of $K(\theta)$, then $\theta[(\alpha+\beta)/2] \geqq (\beta-\alpha)^{1/2}$.*

Proof. In order to establish these properties we note first that for each integer n, θ is clearly continuous on $[\lambda_n, \lambda_{n+1}]$ and that

$$\lambda_n \leqq \theta(x) \leqq (\lambda_{n+1})^{1/2}, \qquad \lambda_n \leqq x \leqq \lambda_{n+1}. \tag{1}$$

Thus from (1), the symmetry of θ, and the equalities

$$\lim_{n\to\infty} \lambda_n = \lim_{n\to\infty} (\lambda_n)^{1/2} = 1, \qquad \lim_{n\to-\infty} \lambda_n = \lim_{n\to-\infty} (\lambda_n)^{1/2} = 0,$$

it follows that θ is continuous on $(0, 2)$ with $\theta(\xi)\to 0$ as $\xi\to 0+$. Observing that the relations

$$\lambda_{n+1} - \lambda_n < \lambda_n, \qquad \lambda_n + \lambda_{-n} = 1, \qquad n \text{ an integer},$$

* This is an almost immediate consequence of the existence of non-dense closed sets of positive measure.

† The meanings of $K(\theta)$, $H(\theta)$, and an *interval of* a set are explained in §2.

found in Lemma 3.2 imply the relations

$$\lambda_{n+1} < 2\lambda_n, \qquad \lambda_n = 1 - \lambda_{-n}, \qquad n \text{ an integer}, \tag{2}$$

and recalling that the geometric mean never exceeds the arithmetic, we conclude from (1) that $\lambda_n \leqq x \leqq \lambda_{n+1}$ implies

$$0 < x/2 \leqq \lambda_{n+1}/2 < \lambda_n \leqq \theta(x) \leqq (\lambda_{n+1})^{1/2} = (1 - \lambda_{-n-1})^{1/2} \leqq (1 - \lambda_{-n}/2)^{1/2}$$
$$= \left(\frac{1+\lambda_n}{2}\right)^{1/2} = \left[1 \cdot \left(\frac{1+\lambda_n}{2}\right)\right]^{1/2} \leqq \frac{1}{2}\left(1 + \frac{1+\lambda_n}{2}\right)$$
$$= \frac{3+\lambda_n}{4} \leqq \frac{3+x}{4};$$

whence (a) is established.

Using the notation of 3.3 for the remainder of the proof, and deferring consideration of (b) until last, we begin checking (c) by noting that (1) and (2) imply

$$\begin{aligned} |\,\theta(1) - \theta(\xi)\,| &= 1 - \theta(\xi) \leqq 1 - \lambda_n = \lambda_{-n} \leqq 2\lambda_{-n-1} \\ &= 2(1 - \lambda_{n+1}) \leqq 2(1 - \xi), \qquad \lambda_n \leqq \xi \leqq \lambda_{n+1}. \end{aligned}$$

From this and the symmetry of θ it follows that for $x = 1$

$$\lim_{\xi \to x} |\,[\theta(\xi) - \theta(x)]/(\xi - x)\,| < \infty;$$

for other $x \in (0, 2)$, this inequality results from the relation (see 3.3 and note that $|\gamma_{n+1} - \gamma_n| < 1$ for each integer n)

$$|\,\theta(x_2) - \theta(x_1)\,| \leqq |\,x_2 - x_1\,| \,/\, |\,[\lambda_n, \lambda_{n+1}] \cdot \mathfrak{H}\,|, \qquad \lambda_n \leqq x_1, x_2 \leqq \lambda_{n+1}.$$

In order to check (d), we prove first that $\lambda_n \in H(\theta)$ for each integer n. If n is odd, then $\gamma_n = \lambda_n < (\lambda_{n+1})^{1/2} = \gamma_{n+1}$ and $\gamma_{n+1} - \gamma_n > 0$, so that the combination of this last relation with the property of $\mathfrak{H}$ stipulated in 3.3 yields the relations $\lambda_n \in H(\theta)$, $\lambda_{n+1} \in H(\theta)$; whence $\lambda_n \in H(\theta)$ for each integer n. This of course implies $1 \in H(\theta)$, so that upon letting (α, β) be any interval of $K(\theta)$ we have either $(\alpha, \beta) \subset (0, 1)$ or $(\alpha, \beta) \subset (1, 2)$. Thus in order to verify

$$\theta([\alpha + \beta]/2) \geqq (\beta - \alpha)^{1/2}$$

there is, in view of the symmetry of θ, no loss in generality in assuming $(\alpha, \beta) \subset (0, 1)$. Since $\lambda_n \in H(\theta)$ for each integer n, it is clear that an integer N exists for which $\lambda_N \leqq \alpha < \beta \leqq \lambda_{N+1}$, and hence from (1) and Lemma 4.2 follows the relation

$$\theta([\alpha + \beta]/2) \geqq \lambda_N > (\lambda_{N+1} - \lambda_N)^{1/2} \geqq (\beta - \alpha)^{1/2}.$$

The property (d) is now established as well as the second part of (b); the first part of (b) is a consequence of the fact that the complement of $\mathfrak{H}$ is an everywhere dense open set.

4. Functions of class $\mathfrak{A}$; the fundamental operation

4.1. Definition. *A function f on $(-1, 1)$ is said to belong to the class $\mathfrak{A}$ if it possesses the following properties*:

(a) *f is continuous with $0 \leqq f(-x) = f(x)$ for $x \in (-1, 1)$;*

(b) *$K(f)$ and $P(f)$ are dense in $(-1, 1)$;*

(c) *Corresponding to each interval (α, β) of the open set $P(f)$ there exists a number $h \geqq ([\beta-\alpha]/2)^{1/2}$ such that $x \in (\alpha, \beta)$ implies*

$$f(x) = h \cdot \theta[(2x - 2\alpha)/(\beta - \alpha)].$$

In the next definition we exhibit the operation which is fundamental in the present development and which, among other things, associates with each function in $\mathfrak{A}$ another such function.

4.2. Definition. *With each function f on $(-1, 1)$ we associate the function $\bar{f}$, likewise on $(-1, 1)$, for which $x \in H(f)$ implies $\bar{f}(x) = 0$, and $x \in K(f)$ implies $\bar{f}(x) = h \cdot \theta[(2x-2\alpha)/(\beta-\alpha)]$, where (α, β) is the interval of $K(f)$ of which x is an element and h is the lesser of the numbers**

$$\left(\inf \left\{K(f)^{(\beta-\alpha)} \cdot \left[\frac{\alpha+\beta}{2}, 1\right]\right\} - \sup \left\{K(f)^{(\beta-\alpha)} \cdot \left[-1, \frac{\alpha+\beta}{2}\right]\right\}\right)^{1/2}$$

and

$$\frac{|f[(\alpha+\beta)/2]|}{2^{1/2}}.$$

We devote 4.3–4.9 to the study of this operation.

4.3. Lemma. *If $f \in \mathfrak{A}$ and (α, β) is an interval of $K(f)$ with $(\alpha+\beta)/2 = \mu$, then $f(\mu) \geqq (\beta-\alpha)^{1/2}$.*

Proof. The satisfaction of 4.1(b) by f implies $K(f) \subset P(f)$ which in turn implies the existence of an interval (α', β') of $P(f)$ for which $\alpha' < \alpha < \beta < \beta'$. Now let

$$\mu' = \frac{\alpha' + \beta'}{2}, \qquad \alpha'' = \frac{2\alpha - 2\alpha'}{\beta' - \alpha'}, \qquad \beta'' = \frac{2\beta - 2\alpha'}{\beta' - \alpha'}, \qquad \mu'' = \frac{\alpha'' + \beta''}{2},$$

and observe that the relation (see 4.1(c))

* By $K(f)^{(\beta-\alpha)}$ we understand $[K(f)]^{(\beta-\alpha)}$. The latter notation is explained in §2.

$$f(x) = f(\mu')\theta[(2x - 2\alpha')/(\beta' - \alpha')], \qquad \alpha' < x < \beta',$$

implies (α'', β'') is an interval of $K(\theta)$. Making use of 3.4(d), 4.1(c), and the above relations we conclude

$$\begin{aligned} f(\mu) &= f(\mu')\theta[(2\mu - 2\alpha')/(\beta' - \alpha')] = f(\mu')\theta(\mu'') \geqq f(\mu')(\beta'' - \alpha'')^{1/2} \\ &\geqq \{[(\beta' - \alpha')/2]\cdot(\beta'' - \alpha'')\}^{1/2} \\ &= \{[(\beta' - \alpha')/2]\cdot 2(\beta - \alpha)/(\beta' - \alpha')\}^{1/2} = (\beta - \alpha)^{1/2}. \end{aligned}$$

4.4. LEMMA. *If $f \in \mathfrak{A}$ and (α, β) is an interval of $K(f)$ with $(\alpha+\beta)/2 = \mu$, then $\bar{f}(\mu) \geqq [(\beta-\alpha)/2]^{1/2}$ and $x \in (\alpha, \beta)$ implies*

$$0 < \bar{f}(-x) = \bar{f}(x) = \bar{f}(\mu)\theta[(2x - 2\alpha)/(\beta - \alpha)] \leqq \bar{f}(\mu) \leqq f(x)/2^{1/2}.$$

Proof. Letting

$$\alpha_0 = \sup\{K(f)^{(\beta-\alpha)}\cdot[-1, \mu]\}, \quad \beta_0 = \inf\{K(f)^{(\beta-\alpha)}\cdot[\mu, 1]\}$$

so that $\bar{f}(\mu)$ is the lesser of the numbers $f(\mu)/2^{1/2}$ and $(\beta_0-\alpha_0)^{1/2}$, we note that the evenness of f implies $(-\beta, -\alpha)$ is an interval of $K(f)$ and implies the equivalence of the relations $x \in K(f)^{(\beta-\alpha)}$, $-x \in K(f)^{(\beta-\alpha)}$ which in turn imply

$$-\beta_0 = \sup\{K(f)^{(\beta-\alpha)}\cdot[-1, -\mu]\}, \quad -\alpha_0 = \inf\{K(f)^{(\beta-\alpha)}\cdot[-\mu, 1]\}.$$

Consequently, in view of the relation $f(\mu) = f(-\mu)$, it follows that $\bar{f}(\mu) = \bar{f}(-\mu)$, which in conjunction with 4.2 and the symmetry of θ assures us that $\bar{f}(x) = \bar{f}(-x)$ for $x \in (\alpha, \beta)$. Furthermore the relations

$$(\beta_0 - \alpha_0)^{1/2} \geqq (\beta - \alpha)^{1/2} > [(\beta - \alpha)/2]^{1/2}, \quad f(\mu)/2^{1/2} \geqq [(\beta - \alpha)/2]^{1/2},$$

the latter of which is a consequence of 4.3, entail

$$\bar{f}(\mu) \geqq [(\beta - \alpha)/2]^{1/2} > 0.$$

From this last relation, 4.2, and 3.4(a) it follows that $x \in (\alpha, \beta)$ implies

$$0 < \bar{f}(x) = \bar{f}(\mu)\cdot\theta[(2x - 2\alpha)/(\beta - \alpha)] \leqq \bar{f}(\mu) \leqq f(\mu)/2^{1/2} = f(x)/2^{1/2},$$

and the proof of the lemma is complete.

4.5. LEMMA. *If $f \in \mathfrak{A}$ then*

$$0 \leqq \bar{f}(-x) = \bar{f}(x) \leqq f(x)/2^{1/2}, \qquad -1 < x < 1.$$

Proof. Since f is even, $x \in H(f)$ implies $-x \in H(f)$. Hence from 4.2 we conclude that for $x \in H(f)$

$$0 = \bar{f}(-x) = \bar{f}(x) \leqq f(x)/2^{1/2},$$

and the preceding lemma completes the proof.

4.6. LEMMA. *If* $f \in \mathfrak{A}$ *then* $K(f)=P(\bar{f})$, $H(f)=Z(\bar{f})$.

Proof. Lemma 4.4 implies $K(f) \subset P(\bar{f})$. Definition 4.2 implies $H(f) \subset Z(\bar{f})$. Lemma 4.5 implies that $P(\bar{f})$ and $Z(\bar{f})$ are complementary in $(-1, 1)$. The proof is complete.

4.7. LEMMA. *If* $f \in \mathfrak{A}$ *and* (α, β) *is an interval of* $K(f)$, *then* α *and* $(\alpha+\beta)/2$ *are elements of* $H(\bar{f})$ *and we have*

$$\bar{f}(\alpha) = \lim_{\xi\to\alpha+} \bar{f}(\xi) = 0 \leqq \liminf_{\substack{\xi\to\alpha+ \\ \xi\in H(\bar{f})}} \frac{\bar{f}(\xi)-\bar{f}(\alpha)}{\xi-\alpha} < \limsup_{\substack{\xi\to\alpha+ \\ \xi\in H(\bar{f})}} \frac{\bar{f}(\xi)-\bar{f}(\alpha)}{\xi-\alpha} = \infty .$$

Proof. Referring to 4.4 we note

$$\bar{f}(\xi) = h\theta[(2\xi-2\alpha)/(\beta-\alpha)], \qquad \alpha < \xi < \beta,$$

where $h=\bar{f}[(\alpha+\beta)/2]>0$; so that upon noting $\alpha \in H(f)$ and referring to 3.4(a), 4.2 we conclude that $\bar{f}(\xi)\to\bar{f}(\alpha)=0$ as $\xi\to\alpha+$. Also upon defining $\xi_n=\alpha+\lambda_n(\beta-\alpha)/2$ and recalling that $\lambda_n \in H(\theta)$ for each integer n we conclude that $\xi_n \in H(\bar{f})$ for each integer n and (see 3.3)

$$\frac{\bar{f}(\xi_n)}{\xi_n-\alpha} = \frac{h\theta(\lambda_n)}{\lambda_n}\cdot\frac{2}{\beta-\alpha} = \frac{\lambda_n}{\lambda_n}\cdot\frac{2h}{\beta-\alpha} \to \frac{2h}{\beta-\alpha}, \qquad n \text{ odd}, n\to-\infty,$$

$$\frac{\bar{f}(\xi_n)}{\xi_n-\alpha} = \frac{h\theta(\lambda_n)}{\lambda_n}\cdot\frac{2}{\beta-\alpha} = \frac{(\lambda_n)^{1/2}}{\lambda_n}\cdot\frac{2h}{\beta-\alpha} \to \infty, \qquad n \text{ even}, n\to-\infty.$$

Since $\xi_n\to\alpha$ or $(\alpha+\beta)/2$ according as $n\to-\infty$ or $+\infty$, it is clear that α and $(\alpha+\beta)/2$ are elements of $H(\bar{f})$ and that the proof of the lemma is complete.

4.8. THEOREM. *If* $f \in \mathfrak{A}$ *then* $x \in H(f)\cdot P(f)$ *implies*

$$\lim_{\xi\to x+} \bar{f}(\xi) = \bar{f}(x) = 0 \leqq \liminf_{\substack{\xi\to x+ \\ \xi\in H(\bar{f})}} \frac{\bar{f}(\xi)-\bar{f}(x)}{\xi-x} < \limsup_{\substack{\xi\to x+ \\ \xi\in H(\bar{f})}} \frac{\bar{f}(\xi)-\bar{f}(x)}{\xi-x} = \infty .$$

Proof. In view of the preceding lemma we confine ourselves to the case in which x not only is an element of $H(f)\cdot P(f)$ but also a cluster point of $[x, 1]\cdot H(f)$.

Let $M(r)=\sup_{-1<t<1}\{\inf K(f)^{(r)}\cdot[t, 1]-\sup K(f)^{(r)}\cdot[-1, t]\}$, and note that the openness of $K(f)$ implies $(-1, 1)\cdot K(f)^{(r)} \uparrow K(f)$ as $r\to0+$ so that the openness and denseness of $K(f)$ in $(-1, 1)$ imply* $M(r)\to0$ as $r\to0+$. Thus if $\{\xi_n\}$ is a sequence of points of $K(f)$ tending to x from the right and if the

* Let $\epsilon>0$; let $-1=a_0<a_1<\cdots<a_N=1$ with $a_\nu-a_{\nu-1}<\epsilon/2$ for $\nu=1, 2, \cdots, N$. Since there exists a positive number δ such that $K(f)^{(\delta)}\cdot[a_{\nu-1}, a_\nu]$ is non-void for $\nu=1, 2, \cdots, N$, we conclude that $0<r<\delta$ implies $M(r)<\epsilon$.

interval of $K(f)$ of which ξ_n is an element is denoted by (α_n, β_n), it becomes clear in the light of 4.5, 4.4, and 4.2 that

$$0 = \bar{f}(x) \leqq \bar{f}(\xi_n) \leqq \bar{f}[(\alpha_n + \beta_n)/2] \leqq [M(\beta_n - \alpha_n)]^{1/2} \to 0 \text{ as } n \to \infty .$$

Since $\bar{f}(x) = \bar{f}(\xi) = 0$ for $\xi \in H(f)$, it is now evident that $\bar{f}(\xi) \to \bar{f}(x) = 0$ as $\xi \to x+$. Now a sequence $\{(a_n, b_n)\}$ of intervals of $K(f)$ with $m_n = (a_n + b_n)/2$ exists such that $m_n \to x+$ and such that for each positive integer n the set $[x, m_n] \cdot K(f)^{(b_n - a_n)}$ is void and* the number $[2M(b_n - a_n)]^{1/2}$ is less than $f(m_n)$. Recalling 4.2 we have

$$\bar{f}(m_n) \leqq [M(b_n - a_n)]^{1/2} < f(m_n)/2^{1/2}, \quad n = 1, 2, 3, \cdots ,$$

which with the fact that $[x, m_n] \cdot K(f)^{(b_n - a_n)}$ is void for $n = 1, 2, 3, \cdots$ implies

$$\begin{aligned} \bar{f}(m_n) &= \{\inf [m_n, 1] \cdot K(f)^{(b_n - a_n)} - \sup [-1, m_n] \cdot K(f)^{(b_n - a_n)}\}^{1/2} \\ &\geqq (m_n - x)^{1/2}, \qquad n = 1, 2, 3, \cdots . \end{aligned}$$

The Theorem is now a consequence of 4.5, the first conclusion of 4.7, and the following two relations which are valid for $n = 1, 2, 3, \cdots$:

$$\bar{f}(m_n)/(m_n - x) \geqq 1/(m_n - x)^{1/2}, \qquad \bar{f}(a_n) = \bar{f}(x) = 0.$$

4.9. THEOREM. *If $f \in \mathfrak{A}$ then $\bar{f} \in \mathfrak{A}$.*

Proof. That $\bar{f}$ is continuous on the right at each point of $H(f) \cdot P(f)$ is a consequence of 4.8; that the same is true on $H(f) \cdot Z(f)$ and on $K(f)$ is a consequence of 4.5 in the first case, and of 4.4, 3.4(a) in the second; that $\bar{f}$ satisfies 4.1(a) is now a consequence of 4.5.

Since $K(\theta)$ is dense in $(0, 2)$ it follows from 4.4 that $K(\bar{f})$ is dense in $K(f)$ which is dense in $(-1, 1)$. Lemma 4.6 implies $K(f) = P(\bar{f})$ which of course assures us that $P(\bar{f})$ is dense in $(-1, 1)$ and that $\bar{f}$ exhibits property 4.1(b).

Lemma 4.4 combines with 4.6 to yield the satisfaction by $\bar{f}$ of 4.1(c).

The proof is now complete.

5. THE FUNCTION F

5.1. DEFINITION. *$\{F_n\}$ is the sequence of functions on $(-1, 1)$ determined by*

$$F_0(x) = \theta(x + 1) \text{ for } -1 < x < 1, \qquad F_{n+1} = \bar{F}_n \text{ for } n = 0, 1, 2, \cdots .$$

F is the function defined by the relation

$$F = \sum_{n=0}^{\infty} (-1)^n F_n .$$

* Here we use the continuity of f, the inequality $f(x) > 0$, and the fact that $M(r) \to 0$ as $r \to 0+$.

We are now prepared to establish the following theorem:

5.2. THEOREM. *F is even and continuous on $(-1, 1)$ with*

$$\liminf_{\xi \to x+} \left| \frac{F(\xi) - F(x)}{\xi - x} \right| < \limsup_{\xi \to x+} \left| \frac{F(\xi) - F(x)}{\xi - x} \right| = \infty, \qquad -1 < x < 1,$$

*and, in particular,**

$$\liminf_{\xi \to x+} \left| \frac{F(\xi) - F(x)}{\xi - x} \right| = 0, \qquad x \in \prod_{\nu=0}^{\infty} P(F_\nu).$$

Proof. Denoting $\sum_{\nu=0}^{n}(-1)^\nu F_\nu$ by S_n, $F - S_n$ by R_n, $P(F_n)$ by $\mathfrak{P}_n$, $Z(F_n)$ by $\mathfrak{Z}_n$, $\prod_{\nu=0}^{\infty}\mathfrak{P}_\nu$ by $\mathfrak{P}$, $\sum_{\nu=0}^{\infty}\mathfrak{Z}_\nu$ by $\mathfrak{Z}$, and checking that $F_0 \in \mathfrak{A}$, we deduce from 4.9 that $n = 0, 1, 2, \cdots$ implies $F_n \in \mathfrak{A}$ and hence that F_n is even and continuous. That F is even is now obvious; that it is continuous is a consequence of the easily verified relations (see 4.5)

$$(1) \qquad |R_n(x)| \leqq F_{n+1}(x) \leqq F_n(x)/2^{1/2} \leqq F_0(x)/2^{(n+1)/2} \leqq 2^{-(n+1)/2}, \qquad -1 < x < 1;\ n = 0, 1, 2, \cdots.$$

A consequence of (1) of which considerable implicit use will be made is the fact that $F_{n+1}(x) = 0$, or $x \in \mathfrak{Z}_{n+1}$, implies $R_n(x) = 0$. It also follows from (1) that $\mathfrak{P}_n$ and $\mathfrak{Z}_n$ are complementary sets in $(-1, 1)$ for $n = 0, 1, 2, \cdots$, that $\mathfrak{P}$ and $\mathfrak{Z}$ are likewise complementary in $(-1, 1)$, and that $\mathfrak{Z}_0 \subset \mathfrak{Z}_1 \subset \cdots \subset \mathfrak{Z}_m \subset \cdots$ with $(-1, 1) = \mathfrak{P}_0 \supset \mathfrak{P}_1 \supset \cdots \supset \mathfrak{P}_m \supset \cdots$.

We divide the remainder of the proof into three parts.

Part I. *If $x \in \mathfrak{Z}$ then*

$$\liminf_{\xi \to x+} \left| \frac{F(\xi) - F(x)}{\xi - x} \right| < \limsup_{\xi \to x+} \left| \frac{F(\xi) - F(x)}{\xi - x} \right| = \infty.$$

Letting x_0 be any element of $\mathfrak{Z}$ we introduce the notation

$$\Delta f(\xi) = [f(\xi) - f(x_0)]/(\xi - x_0)$$

to facilitate verification of $\liminf_{\xi \to x_0+} |\Delta F(\xi)| < \limsup_{\xi \to x_0+} |\Delta F(\xi)| = \infty$. To this end let N $(\geqq 1)$ be the integer for which $x_0 \in \mathfrak{Z}_N \cdot \mathfrak{P}_{N-1}$. Then since $x_0 \in \mathfrak{P}_n$ for $0 \leqq n \leqq N-1$, it follows from 4.1(c) and 3.4(c) that for $0 \leqq n \leqq N-1$, $\limsup_{\xi \to x_0+} |\Delta F_n(\xi)| < \infty$ and hence $\limsup_{\xi \to x_0+} |\Delta S_{N-1}(\xi)| < \infty$. On the other hand $\Delta R_N(\xi) = 0$ for $\xi \in \mathfrak{Z}_{N+1}$. Moreover, in view of the relations $F_{N-1} \in \mathfrak{A}$, $F_N = \overline{F}_{N-1}$, and (see 4.6)

$$H(F_n) = Z(F_{n+1}), \qquad n = 0, 1, 2, \cdots,$$

* The set $\prod_{\nu=0}^{\infty} P(F_\nu)$ is a residual set of measure zero. The construction can be modified so that this set has measure arbitrarily close to two.

it may be seen that use of 4.8 yields

$$0 \leqq \liminf_{\substack{\xi \to x_0+ \\ \xi \in \mathfrak{Z}_{N+1}}} \Delta F_N(\xi) < \limsup_{\substack{\xi \to x_0+ \\ \xi \in \mathfrak{Z}_{N+1}}} \Delta F_N(\xi) = \infty .$$

Thus from the above follow the two relations

$$\begin{aligned} \liminf_{\xi \to x_0+} |\Delta F(\xi)| &\leqq \liminf_{\substack{\xi \to x_0+ \\ \xi \in \mathfrak{Z}_{N+1}}} |\Delta F(\xi)| = \liminf_{\substack{\xi \to x_0+ \\ \xi \in \mathfrak{Z}_{N+1}}} |\Delta S_{N-1}(\xi) + \Delta F_N(\xi)| \\ &\leqq \limsup_{\substack{\xi \to x_0+ \\ \xi \in \mathfrak{Z}_{N+1}}} |\Delta S_{N-1}(\xi)| + \liminf_{\substack{\xi \to x_0+ \\ \xi \in \mathfrak{Z}_{N+1}}} |\Delta F_N(\xi)| < \infty \end{aligned}$$

and

$$\begin{aligned} \limsup_{\xi \to x_0+} |\Delta F(\xi)| &\geqq \limsup_{\substack{\xi \to x_0+ \\ \xi \in \mathfrak{Z}_{N+1}}} |\Delta S_{N-1}(\xi) + \Delta F_N(\xi)| \\ &\geqq \limsup_{\substack{\xi \to x_0+ \\ \xi \in \mathfrak{Z}_{N+1}}} |\Delta F_N(\xi)| - \limsup_{\substack{\xi \to x_0+ \\ \xi \in \mathfrak{Z}_{N+1}}} |\Delta S_{N-1}(\xi)| = \infty , \end{aligned}$$

which in conjunction with the fact that x_0 is any point of $\mathfrak{Z}$ complete the proof of Part I.

Part II. *If* $x \in \mathfrak{P}$, *then*

$$\liminf_{\xi \to x+} |[F(\xi) - F(x)]/(\xi - x)| = 0.$$

Let x_0 be any point of $\mathfrak{P}$, let (α_n, β_n) be the interval of $\mathfrak{P}_n$ of which x_0 is an element, let $\mu_n = (\alpha_n + \beta_n)/2$. From 4.6 follows

$$(\alpha_n, \beta_n) \subset \mathfrak{P}_n \subset \mathfrak{P}_{\nu+1} = P(F_{\nu+1}) = K(F_\nu), \qquad 0 \leqq \nu \leqq n-1,$$

which implies

$$\begin{aligned} &F_\nu(\mu_n) - F_\nu(x_0) = F_\nu(\beta_n) - F_\nu(x_0) = 0, \\ &S_\nu(\mu_n) - S_\nu(x_0) = S_\nu(\beta_n) - S_\nu(x_0) = 0, \qquad 0 \leqq \nu \leqq n-1. \end{aligned} \tag{2}$$

Thus for n even and positive (note that $\beta_n \in \mathfrak{Z}_n$, and see inequality (1))

$$\begin{aligned} F(\beta_n) - F(x_0) &= S_{n-1}(\beta_n) - S_{n-1}(x_0) + F_n(\beta_n) - F_n(x_0) + R_n(\beta_n) - R_n(x_0) \\ &= -F_n(x_0) - R_n(x_0) \leqq -F_n(x_0) + |R_n(x_0)| \\ &\leqq -F_n(x_0) + F_n(x_0)/2^{1/2} = (-1 + 2^{-1/2})F_n(x_0) \leqq 0, \end{aligned} \tag{3}$$

whereas for n odd and positive

$$F(\beta_n) - F(x_0) = F_n(x_0) - R_n(x_0) \geqq F_n(x_0) - |R_n(x_0)| \geqq F_n(x_0)(1 - 2^{-1/2}) \geqq 0.$$

These last two relations coupled with the continuity of F imply the existence of $\{\xi_n\}$ for which

$$F(\xi_n) - F(x_0) = 0, \qquad \beta_{n+1} \leqq \xi_n \leqq \beta_n, \qquad n = 1, 2, 3, \cdots .$$

Hence in view of the relation (valid for $n = 1, 2, 3, \cdots$)

$$(4) \qquad 0 < \xi_n - x_0 \leqq \beta_n - x_0 \leqq \beta_n - \alpha_n \leqq 2[F_n(\mu_n)]^2 \leqq 2^{-n+1} \to 0 \text{ as } n \to \infty$$

which is a consequence of 4.1(c), the equality $\theta(1) = 1$, and inequality (1), the remainder of the proof of Part II is evident.

Part III. *If* $x \in \mathfrak{P}$ *then*

$$\limsup_{\xi \to x+} \left| [F(\xi) - F(x)]/(\xi - x) \right| = \infty .$$

Let x_0 be any element of $\mathfrak{P}$, and let α_n, β_n, and μ_n be as defined in Part II. By reasoning to be given later we have, in case $\alpha_n < x_0 < \mu_n$ with n even and positive,

$$\begin{aligned} |F(\mu_n) - F(x_0)| &= |F_n(\mu_n) - F_n(x_0) - R_n(x_0)| \geqq F_n(\mu_n) - F_n(x_0) - R_n(x_0) \\ &\geqq F_n(\mu_n) - F_n(x_0) = F_n(\mu_n)\left\{1 - \theta\left(\frac{2x_0 - 2\alpha_n}{\beta_n - \alpha_n}\right)\right\} \\ (5) \qquad &\geqq F_n(\mu_n)\left\{1 - \frac{1}{4}\left(\frac{2x_0 - 2\alpha_n}{\beta_n - \alpha_n} + 3\right)\right\} = F_n(\mu_n)\cdot\left(\frac{\mu_n - x_0}{2\beta_n - 2\alpha_n}\right) \\ &\geqq \left(\frac{\beta_n - \alpha_n}{2}\right)^{1/2}\cdot\left(\frac{\mu_n - x_0}{2\beta_n - 2\alpha_n}\right) = \frac{\mu_n - x_0}{2^{3/2}(\beta_n - \alpha_n)^{1/2}} \\ &\geqq 2^{(n-4)/2}\cdot(\mu_n - x_0) \geqq 2^{(n-6)/2}(\mu_n - x_0), \end{aligned}$$

whereas in case $\mu_n \leqq x_0 < \beta_n$ with n even and positive we have

$$\begin{aligned} |F(\beta_n) - F(x_0)| &= |F_n(x_0) + R_n(x_0)| \geqq F_n(x_0) - |R_n(x_0)| \\ &\geqq F_n(x_0)(1 - 2^{-1/2}) \\ &\geqq 2^{-2}F_n(x_0) = 2^{-2}F_n(\mu_n)\cdot\theta\left(\frac{2x_0 - 2\alpha_n}{\beta_n - \alpha_n}\right) \\ (6) \qquad &= 2^{-2}F_n(\mu_n)\cdot\theta\left(\frac{2\beta_n - 2x_0}{\beta_n - \alpha_n}\right) \geqq 2^{-2}F_n(\mu_n)\left(\frac{\beta_n - x_0}{\beta_n - \alpha_n}\right) \\ &\geqq 2^{-2}\left(\frac{\beta_n - \alpha_n}{2}\right)^{1/2}\left(\frac{\beta_n - x_0}{\beta_n - \alpha_n}\right) \geqq 2^{-5/2}\frac{(\beta_n - x_0)}{(\beta_n - \alpha_n)^{1/2}} \\ &\geqq 2^{(n-6)/2}\cdot(\beta_n - x_0). \end{aligned}$$

Consequently $\{\xi_n\}$ exists for which $x_0 < \xi_n = \mu_{2n}$ or β_{2n}, $(n = 1, 2, 3, \cdots)$, with the result that

$$\left|\frac{F(\xi_n) - F(x_0)}{\xi_n - x_0}\right| \geqq 2^{n-3} \to \infty \quad \text{as} \quad n \to \infty$$

and the proof of the theorem will be complete as soon as we verify (5) and (6).

We observe that (5) consists of ten steps the reasons for which we now give in order. (i) This is a consequence of relation (2) of Part II and the fact that 4.7 and 4.6 imply $\mu_n \in H(F_n) = Z(F_{n+1}) = \mathfrak{Z}_{n+1}$. (ii) Arithmetic. (iii) In an alternating series of the type considered here the remainder term has the same sign as the first term omitted. Thus $(-1)^{n+1}F_{n+1}(x_0)$ and $R_n(x_0)$ are non-positive since n is even and positive. (iv) This is a consequence of 4.1(c) and $\theta(1) = 1$. (v) From 3.4(a) it follows that $\theta(x) \leqq (x+3)/4$ for $0 < x < 1$. (vi) Arithmetic and the definition of μ_n. (vii) Inequality (4) of Part II. (viii) Arithmetic. (ix) Inequality (4) of Part II. (x) Arithmetic.

We treat (6) in a similar manner. (i) Relation (3) of Part II. (ii) An elementary inequality. (iii) Relation (1). (iv) Arithmetic. (v) This is a consequence of 4.1(c) and $\theta(1) = 1$. (vi) From 3.3 we have $\theta(x) = \theta(2-x)$ for $0 < x < 2$. (vii) From 3.4(a) we have $x/2 \leqq \theta(x)$ for $0 < x < 1$. (viii) Inequality (4) of Part II. (ix) Arithmetic. (x) Inequality (4) of Part II.

THE INSTITUTE FOR ADVANCED STUDY,
PRINCETON, N. J.

A SYSTEM OF ORDINARY LINEAR DIFFERENTIAL EQUATIONS WITH TWO-POINT BOUNDARY CONDITIONS*

BY

WILLIAM T. REID

1. **Introduction.** In a recent paper Bliss [2]† has treated a boundary value problem which is definitely self-adjoint according to a weakened form of an earlier definition of definite self-adjointness which he gave in a paper [2] published in 1926. He has shown that a system which satisfies this modified definition has most of the properties proved in the former paper for a system satisfying the original and stronger definition of definite self-adjointness. The characteristic values for such a problem are all real and have indices equal to their multiplicities, and the expansion theorems established in the earlier paper are still valid. Such a system, however, will not in general have an infinity of characteristic values.

Bliss has shown that the canonical form of the so-called accessory boundary value problem for a normal non-singular problem of Bolza in the calculus of variations is definitely self-adjoint according to this new definition, while such a problem is in general not definitely self-adjoint according to the original definition. Hu [5] has proved that this problem has an infinity of characteristic values. Various authors (Morse [7] and [8], Reid [9], [10], and [11], Hu [5], Hölder [4], Birkhoff and Hestenes [1], Wiggin [12]) have treated the accessory problem under the assumptions that the problem is normal and satisfies the strengthened Clebsch condition. This latter condition implies non-singularity. Under these hypotheses the characteristic values possess certain extremizing properties. Moreover, for such systems there have been established oscillation and comparison theorems which are generalizations of the classic results of the Sturmian theory for a single second order differential equation (see, for example, Ince [6], chap. 10). It is to be remarked that for the problem treated by the author ([9] and [10]) it is not assumed that the coefficients of the terms involving the parameter are those of a definite quadratic form. To compensate for this weakened condition it it assumed that the quadratic functional of the associated minimum problem is definite. This latter hypothesis is no additional restriction for an accessory problem which is normal, satisfies the strengthened Clebsch condition, and

* Presented to the Society, April 11, 1936; received by the editors November 3, 1937.

† Numerals in square brackets refer to the bibliography at the end of the present paper.

for which the coefficients of the terms involving the parameter belong to a definite quadratic form. In the author's paper [9] there are also established expansion theorems in terms of the characteristic solutions of the problem.

It is the purpose of the present paper to consider a boundary problem which is definitely self-adjoint according to the new definition of Bliss, and which satisfies the additional condition that the matrix of coefficients of the terms involving the parameter has constant rank on the interval of definition. This latter condition is satisfied in the most important examples of definitely self-adjoint systems. It is here shown that such a system is equivalent to a boundary value problem associated with the second variation of a calculus of variations problem, and of the type previously treated by the author [9]. The character of the equivalence is new in the sense that the canonical form of the second problem is not the given definitely self-adjoint problem, but rather a system involving twice the number of dependent functions occurring in the original system. There is obtained a general condition which is both necessary and sufficient for the given system to have an infinity of characteristic values. There are also given two different sets of conditions which are merely sufficient for this conclusion. In particular, in §5 it is shown that an accessory boundary value problem which is normal and non-singular has an infinity of characteristic values. As stated above, this result was first proved by Hu. As already noted, oscillation, comparison, and expansion theorems are known for a system of the form to which the given definitely self-adjoint problem is here shown to be equivalent. From these theorems there are deduced in §6 certain corresponding results for the given boundary value problem.

The notation used for the definitely self-adjoint system here treated is the same as that used by Bliss in papers [2] and [3]. Frequent references are made to the results concerning differential systems in general, and the theorems for self-adjoint systems, obtained by Bliss in §§1 and 2 of paper [2]. It is to be noted that for the particular definitely self-adjoint problem herein treated the proof of the reality of characteristic values is independent of the proof of paper [3] of Bliss, since this result is here deduced from the corresponding theorem for the equivalent problem of the type previously treated by the author.

2. **Statement of problem.** In the following pages the summation convention of tensor analysis will be used. The subscripts i, j, k will have the range $1, \cdots, n$. The matrices $\|A_{ij}(x)\|$, $\|B_{ij}(x)\|$ will be supposed to have elements which are real-valued and continuous on $a \leqq x \leqq b$, while the elements of the matrices $\|M_{ij}\|$, $\|N_{ij}\|$ are real constants such that the matrix $\|M_{ij}; N_{ij}\|$ has rank n. For brevity we will introduce the notations*

* Here, as elsewhere, primes are used to denote derivatives with respect to the variable x.

$$(2.1)\qquad \mathcal{L}_i[y] = y_i' - A_{ij}(x)y_j, \qquad \mathcal{M}_i[z] = z_i' + z_jA_{ji}(x).$$

The boundary value problem to be considered in this paper may then be written as

$$(2.2)\qquad \mathcal{L}_i[y] = \lambda B_{ij}(x)y_j, \quad s_i[y] \equiv M_{ij}y_j(a) + N_{ij}y_j(b) = 0.$$

The system adjoint to (2.2) is (see Bliss [2]),

$$(2.3)\qquad \mathcal{M}_i[z] = -\lambda z_jB_{ji}(x), \quad t_i[z] = z_j(a)P_{ji} + z_j(b)Q_{ji} = 0,$$

where $p_i = P_{ij}$, $q_i = Q_{ij}$, $(j = 1, \cdots, n)$, are linearly independent solutions of the algebraic equations

$$M_{ij}p_j - N_{ij}q_j = 0.$$

The boundary value problem (2.2) will be treated under the following hypotheses:

(H_1) The system (2.2) is self-adjoint under the non-singular transformation $z_i = T_{ij}(x)y_j$, that is, under this transformation the differential equations and boundary conditions of the adjoint system (2.3) are equivalent to those of (2.2) for all values of λ. The elements $T_{ij}(x)$ are supposed to be of class C^1 on the interval ab.

A non-singular matrix $\|T_{ij}(x)\|$ of functions of class C^1 is such a transformation matrix if and only if

$$(2.4)\qquad \begin{gathered} T_{ik}A_{kj} + A_{ki}T_{kj} + T_{ij}' \equiv 0, \qquad T_{ik}B_{kj} + B_{ki}T_{kj} \equiv 0, \\ M_{ik}T_{kl}^{-1}(a)M_{jl} = N_{ik}T_{kl}^{-1}(b)N_{jl}. \end{gathered}$$

In particular, if $u_i(x)$ and $v_i(x)$ are sets of functions of class C^1 and are related by the equations $v_i(x) \equiv T_{ji}(x)u_j(x)$, we have the identity

$$(2.5)\qquad \mathcal{M}_i[v] \equiv T_{ji}(x)\mathcal{L}_j[u].$$

Moreover, $t_i[v] = 0$ if and only if $s_i[u] = 0$.

(H_2) $\|S_{ij}(x)\| \equiv \|T_{ki}(x)B_{kj}(x)\|$ is symmetric.

(H_3) If (u_i) is a set of real constants, then

$$u_iS_{ij}(x)u_j \geqq 0.$$

(H_4) There exists no non-identically vanishing solution of $\mathcal{L}_i[y] = 0$, $s_i[y] = 0$ such that $B_{ij}(x)y_j(x) \equiv 0$ on ab.

If the system (2.2) is the canonical form of a so-called accessory boundary value problem associated with the second variation of a problem of the calculus of variations, the above hypothesis (H_4) is equivalent to the assumption

of normality of the accessory problem. The significance of this hypothesis for the general self-adjoint system of the form (2.2) has been pointed out by Bliss [3].

(H_5) $\|B_{ij}(x)\|$ is of constant rank $n-m$, $(0 \leqq m < n)$, on $a \leqq x \leqq b$.

If $y_i(a)$, $y_i(b)$, $z_i(a)$, $z_i(b)$ are arbitrary values, we have (Bliss [2]),

$$y_i(x)z_i(x)\Big|_a^b = \bar{t}_i[z]s_i[y] + t_i[z]\bar{s}_i[y], \tag{2.6}$$

where $\bar{s}_i[y] = \overline{M}_{ij}y_j(a) + \overline{N}_{ij}y_j(b)$, $\bar{t}_i[z] = z_j(a)\overline{P}_{ji} + z_j(b)\overline{Q}_{ji}$, and the matrices $\|\overline{M}_{ij}\|$, $\|\overline{N}_{ij}\|$, $\|\overline{P}_{ij}\|$, $\|\overline{Q}_{ij}\|$ are chosen so that

$$\left\| \begin{matrix} M_{ij} & N_{ij} \\ \overline{M}_{ij} & \overline{N}_{ij} \end{matrix} \right\|, \qquad \left\| \begin{matrix} -\overline{P}_{ij} & -P_{ij} \\ \overline{Q}_{ij} & Q_{ij} \end{matrix} \right\|$$

are reciprocals. Moreover, if $y_i(x)$, $z_i(x)$ are solutions of the differential equations of the systems (2.2) and (2.3), respectively, for some value of λ, then $[y_i(x)z_i(x)]' = 0$. Combining these results, and making use of the self-adjointness of the system (2.2) under the transformation $z_i = T_{ij}y_j$, we obtain that if $y_i(x)$ and $y_i^*(x)$ are solutions of (2.2) corresponding to distinct values λ and λ^*, respectively, then (see Bliss [2])

$$\int_a^b y_i(x)S_{ij}(x)y_j^*(x)dx = 0. \tag{2.7}$$

3. **Relation of the boundary value problem to the calculus of variations.** In this section there will be pointed out the relation of the above boundary value problem (2.2) with a problem of the form of an accessory boundary problem associated with the calculus of variations.

As a consequence of hypotheses (H_2), (H_3), and (H_5) there exist m sets of functions $\pi_{i\alpha}(x)$, $(\alpha = 1, \cdots, m)$, such that

$$S_{ij}(x)\pi_{j\alpha}(x) \equiv 0, \qquad i = 1, \cdots, n;\ \alpha = 1, \cdots, m;\ a \leqq x \leqq b. \tag{3.1}$$

Moreover, these functions $\pi_{i\alpha}(x)$ may be chosen as continuous on $a \leqq x \leqq b$, and orthonormal in the sense that

$$\pi_{i\alpha}(x)\pi_{i\beta}(x) = \delta_{\alpha\beta}, \qquad \alpha, \beta = 1, \cdots, m. \tag{3.2}$$

In view of (H_3) it follows that the matrix

$$\left\| \begin{matrix} S_{ij}(x) & \pi_{i\beta}(x) \\ \pi_{j\alpha}(x) & 0_{\alpha\beta} \end{matrix} \right\|$$

is non-singular on ab. The reciprocal of this matrix is seen to be of the form

$$\left\| \begin{matrix} R_{ij}(x) & \pi_{i\beta}(x) \\ \pi_{j\alpha}(x) & 0_{\alpha\beta} \end{matrix} \right\|, \tag{3.3}$$

and $R_{ij}(x)u_iu_j>0$ for every set $(u_i)\neq(0_i)$ satisfying the relations $\pi_{j\alpha}(x)u_j=0$, $(\alpha=1,\cdots,m)$.

Finally, we shall suppose that there are exactly ρ linearly independent solutions $y_i=u_{i\kappa}(x)$, $(\kappa=1,\cdots,\rho)$, of the system (2.2) for $\lambda=0$. Bliss has shown that under the hypotheses (H_1)–(H_4) the index of a characteristic value is equal to its multiplicity as a root of the characteristic equation $D(\lambda)=0$. Hence not all values of λ can be characteristic values of (2.2) under these hypotheses, and without loss of generality we might assume $\lambda=0$ is not a characteristic value. The method of proof as given by Bliss is similar to that used in his previous paper [2] to establish the same result under stronger hypotheses. However, for completeness and independence of the results of the present paper, use will not be made of this result. It is to be pointed out that Theorem 3.1 below includes the result that under hypotheses (H_1)–(H_5) all values of λ cannot be characteristic values of (2.2).

An arc $z\equiv[z_i(x)]$ will be called admissible if the functions $z_i(x)$ are of class D' on ab, and satisfy the differential equations

$$\pi_{j\alpha}(x)\mathfrak{M}_j[z]=0, \qquad \alpha=1,\cdots,m. \tag{3.4}$$

For brevity, the class H_1^* will be defined as the totality of admissible arcs z satisfying the conditions

$$t_i[z]\equiv z_j(a)P_{ji}+z_j(b)Q_{ji}=0, \qquad i=1,\cdots,n, \tag{3.5}$$

$$\int_a^b z_i(x)B_{ij}(x)u_{j\kappa}(x)dx=0, \qquad \kappa=1,\cdots,\rho, \tag{3.6}$$

$$\int_a^b z_i(x)K_{ij}(x)z_j(x)dx=1, \tag{3.7}$$

where $K_{ij}(x)=-B_{ik}(x)T_{jk}^{-1}(x)$. Since $K_{ij}(x)=T_{ik}^{-1}(x)S_{kl}(x)T_{jl}^{-1}(x)$, the matrix $\|K_{ij}(x)\|$ is symmetric and $u_iK_{ij}(x)u_j\geqq 0$ for all real sets $(u_i)\neq(0_i)$.

Suppose the class H_1^* is non-vacuous, and consider the problem of minimizing the integral

$$I[z]=\int_a^b \mathfrak{M}_i[z]R_{ij}(x)\mathfrak{M}_j[z]dx \tag{3.8}$$

in this class. In view of (3.4) we have that $I[z]$ is non-negative in the class of admissible arcs.

Because of the integral conditions (3.6) this minimum problem is not of

the form of an accessory problem associated with the problem of Bolza in the calculus of variations. It may be reduced, however, to such a problem by the introduction of new variables $z_{n+\kappa}$ satisfying the differential equations $z'_{n+\kappa} - z_i B_{ij} u_{j\kappa} = 0$ and end conditions $z_{n+\kappa}(a) - z_{n+\kappa}(b) = 0$.

If, for brevity, we set

$$R_{ij}(x)\mathfrak{M}_j[z] + \pi_{i\alpha}(x)\mu_\alpha(x) = \zeta_i(x), \tag{3.9}$$

the first necessary condition for this calculus of variations problem states that there exist multipliers $\mu_\alpha(x)$ and constants σ_κ, Λ, d_i, such that in addition to equations (3.4), (3.5), (3.6), (3.7) we have

$$\zeta_i' - A_{ij}\zeta_j - B_{ij}u_{j\kappa}\sigma_\kappa + \Lambda K_{ij}z_j = 0, \tag{3.10}$$

$$P_{ij}d_j - \zeta_i(a) = Q_{ij}d_j + \zeta_i(b) = 0, \tag{3.11}$$

where $\zeta_i(x)$ is the function defined in terms of $z_i(x)$, $\mu_\alpha(x)$ by (3.9). Solving (3.4) and (3.9) simultaneously, we obtain

$$\mathfrak{M}_i[z] = S_{ij}(x)\zeta_j, \qquad \mu_\alpha(x) = \pi_{j\alpha}(x)\zeta_j.$$

Moreover, since $M_{ik}P_{kj} - N_{ik}Q_{kj} = 0$, the equations (3.11) are equivalent to the conditions $s_i[\zeta] = 0$. Consequently, the system (3.4), (3.5), (3.6), (3.10), (3.11) may be written

$$\mathfrak{M}_i[z] = S_{ij}(x)\zeta_j, \qquad \mathfrak{L}_i[\zeta] = B_{ij}u_{j\kappa}\sigma_\kappa - \Lambda K_{ij}z_j,$$

$$\int_a^b z_i(x)B_{ij}(x)u_{j\kappa}(x)dx = 0, \qquad t_i[z] = 0, \qquad s_i[\zeta] = 0, \tag{3.12}$$

$$i, j = 1, \cdots, n; \kappa = 1, \cdots, \rho.$$

Now, since $y_i = u_{i\kappa}(x)$, $(\kappa = 1, \cdots, \rho)$, is a solution of $\mathfrak{L}_i[y] = 0$, $s_i[y] = 0$, we have that $z_i = v_{i\kappa}(x) = T_{ji}(z)u_{j\kappa}(x)$ is a solution of $\mathfrak{M}_i[z] = 0$, $t_i[z] = 0$. Using this relation, together with (2.6), we obtain in view of (H_3) and (H_4) the following result:

LEMMA 3.1. *If z_i, ζ_i satisfies* (3.12) *with constants σ_κ, Λ, then $\sigma_\kappa = 0$, $(\kappa = 1, \cdots, \rho)$.*

LEMMA 3.2. *The system* (3.12) *is normal.*

For if this system is not normal, there exist functions $z_i(x) \equiv 0$, $\zeta_i(x) \not\equiv 0$ satisfying with constants σ_κ, Λ this system. In view of Lemma 3.2, $\sigma_\kappa = 0$. Then $\zeta_i \equiv \pi_{i\alpha}(x)\mu_\alpha(x)$ is a solution of $\mathfrak{L}_i[\zeta] = 0$, $s_i[\zeta] = 0$. But since $B_{ij}\pi_{j\alpha} \equiv 0$, we then have $B_{ij}(x)\zeta_j \equiv 0$ and by (H_4) we have $\zeta_i(x) \equiv 0$, contrary to the assumption that $\zeta_i(x) \not\equiv 0$. Therefore the system (3.12) is normal.

LEMMA 3.3. *The value $\Lambda = 0$ is not a characteristic value of* (3.12).

This result is a consequence of Lemma 3.1, relation (2.6), hypotheses (H_3) and (H_4), together with the fact that every solution of the system $\mathcal{L}_i[y]=0$, $s_i[y]=0$ is a linear combination of the functions $u_{i\kappa}(x)$, $(\kappa=1, \cdots, \rho)$.

THEOREM 3.1. *The characteristic values* Λ *of* (3.12) *are all real and positive, and at most denumerably infinite in number.*

The reality of the characteristic values has been proved by the author [9]. The positiveness is a consequence of the above hypotheses and the identity

$$I[z] = \Lambda \int_a^b z_i(x)K_{ij}(x)z_j(x)dx$$

which is satisfied by the functions $z_i(x)$ of a solution $z_i, \zeta_i, \sigma_\kappa=0$ of (3.12) for a corresponding value Λ. Since $\Lambda=0$ is not a characteristic value, and the characteristic values are the zeros of a permanently convergent power series (see Bliss [2]), we have at most a denumerable infinity of such values.

The following properties have been proved by Reid [9]:

THEOREM 3.2. *If the class* H_1^* *is non-vacuous and* Λ_1 *is the greatest lower bound of* $I[z]$ *in this class, then* $\Lambda_1>0$ *and* $\Lambda=\Lambda_1$ *is the least characteristic value of* (3.12).

THEOREM 3.3. *Suppose* $\Lambda_1<\Lambda_2<\cdots<\Lambda_{t-1}$ *are consecutive characteristic values of* (3.12), *and corresponding to* $\Lambda=\Lambda_s$, $(s=1, \cdots, t-1)$, *there are* r_s *linearly independent solutions* $z_{iq_s}, \zeta_{iq_s}, \sigma_{\kappa q_s}=0$, $(q_s=1, \cdots, r_s)$. *Define the class* H_t^* *as the subclass of arcs* z *belonging to* H_1^* *and satisfying the additional conditions*

$$0 = \int_a^b z_i(x)K_{ij}(x)z_{jq_s}(x)dx, \qquad q_s = 1, \cdots, r_s; s = 1, \cdots, t-1.$$

If H_t^* *is non-vacuous and* Λ_t *is the greatest lower bound of* $I[z]$ *in this class, then* $\Lambda_t>\Lambda_{t-1}$ *and* $\Lambda=\Lambda_t$ *is a characteristic value of* (3.12).

If $z_i, \zeta_i, \sigma_\kappa=0$ is a solution of (3.12) for a value Λ, set $z_i=(1/\Lambda^{1/2})T_{ji}(x)\eta_j$. Then η_i, ζ_i is a solution of the system

$$(3.13)\quad \mathcal{L}_i[\eta] = \Lambda^{1/2}B_{ij}(x)\zeta_j, \quad \mathcal{L}_i[\zeta] = \Lambda^{1/2}B_{ij}(x)\eta_j, \quad s_i[\eta] = 0, \quad s_i[\zeta] = 0.$$

Now if $\bar\eta_i, \bar\zeta_i$ is a solution of (3.13), the functions $\eta_i=\bar\eta_i+\bar\zeta_i$, $\zeta_i=\bar\eta_i+\zeta_i$ and $\eta_i=\bar\eta_i-\bar\zeta_i$, $\zeta_i=\bar\zeta_i-\bar\eta_i$ are also solutions of this system. Consequently, if the index of Λ as a characteristic value of (3.12) is equal to r, there exist r linearly independent solutions η_i, ζ_i of (3.13) such that for each of these solutions we have either $\eta_i=\zeta_i$ or $\eta_i=-\zeta_i$. Now if $\eta_i, \zeta_i=\eta_i$ is a characteristic solution of (3.13) we see that $y_i=\eta_i$ is a characteristic solution of (2.2) for $\lambda=\Lambda^{1/2}$; simi-

larly, if η_i, $\zeta_i = -\eta_i$ is a characteristic solution of (3.13), we have that $y_i = \eta_i$ is a solution of (2.2) for $\lambda = -\Lambda^{1/2}$. Hence the sum of the indices of $\Lambda^{1/2}$ and $-\Lambda^{1/2}$ as characteristic values of (2.2) is not less than the index of Λ as a characteristic value of (3.12).

Now if $\lambda \neq 0$ is a characteristic value of (2.2) and y_i is a corresponding characteristic solution, then $z_i = (1/\lambda)T_{ji}y_j$, $\zeta_i = y_i$ is a solution of (3.12) for $\Lambda = \lambda^2$. The orthogonality of the functions $z_i(x)$ to the sets $u_{i\kappa}(x)$, $(\kappa = 1, \cdots, \rho)$, is a consequence of (2.7). In view of Lemma 3.1 we have, therefore, that all the characteristic values of (2.2) are real. In view of (2.7) the set of characteristic solutions of (2.2) corresponding to the values $\Lambda^{1/2}$ and $-\Lambda^{1/2}$, where Λ is a given positive value, are linearly independent. Hence the index of Λ as a characteristic value of (3.12) is not less than the sum of the indices of $\Lambda^{1/2}$ and $-\Lambda^{1/2}$ as characteristic values of (2.2). We have, therefore, the following theorem:

THEOREM 3.4. *The characteristic values of* (2.2) *are all real and at most denumerably infinite in number. If* Λ *is a characteristic value of* (3.12), *either* $\Lambda^{1/2}$ *or* $-\Lambda^{1/2}$ *is a characteristic value of* (2.2). *Conversely, if* λ *is a characteristic value of* (2.2), *then* $\Lambda = \lambda^2$ *is a characteristic value of* (3.12) *whose index is equal to the sum of the indices of* λ *and* $-\lambda$ *as characteristic values of* (2.2).

4. **Sufficient conditions for the existence of infinitely many characteristic values.** One may show by simple examples (see Bliss [3]) that the hypotheses of §2 do not imply the existence of infinitely many characteristic values of (2.2). In this section we shall give certain sets of conditions that insure this property. In view of Theorem 3.4 we may assume without loss of generality that $\lambda = 0$ is not a characteristic value of (2.2), and the sets $u_{i\kappa}$, $(\kappa = 1, \cdots, \rho)$, are missing. If this condition is not true for the original problem it may be attained by a linear change of parameter. *We shall make this assumption in the future consideration of this problem.* The following general theorem follows from Theorem 3.4 and the extremizing properties of the characteristic values of (3.12) (see, for example, Reid [9] and [11]).

THEOREM 4.1. *A system* (2.2) *satisfying* (H_1)–(H_5) *has an infinity of characteristic values if and only if there are infinitely many arcs* $z_i = w_{is}(x)$, $(s = 1, 2, \cdots)$, *satisfying* (3.4), (3.5), *and such that for each* r *and arbitrary constants* $(d_t) \neq (0_t)$, $(t = 1, \cdots, r)$, *the arc* $w_i = w_{it}(x)d_t$ *satisfies the condition*

$$\int_a^b w_i K_{ij} w_j dx > 0. \tag{4.1}$$

The following hypothesis is a weakened form of a condition used by Bliss in originally defining definitely self-adjoint systems.

(H_6) If $g_i(x)$ are arbitrary continuous functions on ab, and $y_i(x)$ is a solution of the system

$$\mathcal{L}_i[y] = B_{ij}(x)g_j(x), \qquad s_i[y] = 0, \tag{4.2}$$

then $y_iS_{ij}y_j \equiv 0$ on ab if and only if $y_i \equiv 0$.

THEOREM 4.2. *If* (2.2) *satisfies* (H_1), (H_2), (H_3), (H_5), *and* (H_6), *this system has infinitely many characteristic values.*

Hypotheses (H_2), (H_3), and (H_6) are seen to imply (H_4). Denote by $g_{is}(x)$, $(s=1, 2, \cdots)$, sets of functions such that the sets $B_{ij}(x)g_{js}(x)$ are linearly independent on ab. Such a choice is clearly possible. Now let $y_i = y_{is}(x)$ be a particular solution of (4.2) for $g_i = g_{is}(x)$, and set $w_{is}(x) = T_{ji}(x)y_{js}(x)$, $(s=1, 2, \cdots)$. By the use of relation (2.5) it may be proved that the functions $z_i = w_{is}(x)$ satisfy (3.4) and (3.5). Furthermore, if $(d_t) \neq (0_t)$, $(t=1, \cdots, r)$, and $y_i = y_{it}(x)d_t$, $w_i = w_{it}(x)d_t$ we have

$$\int_a^b w_iK_{ij}w_j dx = \int_a^b y_iS_{ij}y_j dx > 0$$

in view of (H_6). The result of Theorem 4.2 is then a consequence of Theorem 4.1.

COROLLARY. *If the system* (2.2) *satisfies hypotheses* (H_1), (H_2), (H_3), *and* $|B_{ij}| \neq 0$ *on* ab, *then* (2.2) *has infinitely many characteristic values.*

Hypothesis (H_6) is weaker than the condition originally used by Bliss [2] in defining definitely self-adjoint systems since the extra condition $s_i[y]=0$ has been added. The result of Theorem 4.2 is of a somewhat less general character than that originally obtained by Bliss [2], however, since we have assumed (H_5).

THEOREM 4.3. *Suppose that* (2.2) *satisfies* (H_1)–(H_5), *and that the functions* B_{ij} *are of class* C^1 *on* ab. *Then, if the matrix*

$$\|\pi_{i\alpha}(x)\ T_{ki}^{-1}(x)B_{jk}(x)\| \tag{4.3}$$

has at some point x_0 *of* ab *a rank greater than* m, *the system* (2.2) *has infinitely many characteristic values.*

If the functions B_{ij} are of class C^1, the functions $\pi_{i\alpha}(x)$ may also be chosen to be of class C^1, and we shall suppose that these functions are so selected. Now a set z_i satisfies (3.4) if and only if there are functions $g_i(x)$ such that

$$\mathcal{M}_i[z] = -\ g_j(x)B_{ji}(x). \tag{4.4}$$

In view of (H_2) and (H_3) a set of continuous functions $z_i(x)$ renders

$\int_a^b z_i K_{ij} z_j dx = 0$ if and only if the set z_i is of the form $T_{ji}(x)\pi_{j\alpha}(x)f_\alpha(x)$. We shall now determine under what conditions a set of this latter form is a solution of (4.4). In view of (2.5), such a set satisfies (4.4) if and only if

$$\mathcal{L}_i[\pi_\alpha f_\alpha] \equiv \pi_{i\alpha} f_\alpha' + \mathcal{L}_i[\pi_\alpha] f_\alpha = -g_j B_{jk} T_{ki}^{-1}. \tag{4.5}$$

The functions $\mathcal{L}_i[\pi_\alpha]$ are continuous since $\pi_{i\alpha}(x)$ are of class C^1. Now, in view of the hypotheses of the theorem, there is an interval $a_1 b_1$ containing x_0 and integers $1 \leqq i_1 < i_2 < \cdots < i_p \leqq n$, $1 \leqq j_1 < j_2 < \cdots < j_{p-m} \leqq n$, $p > m$, such that the determinant $|\pi_{i_q\alpha} T^{-1}_{i_q l_q} B_{j_h k}|$, $(q = 1, \cdots, p;\ h = 1, \cdots, p-m)$, is different from zero on $a_1 b_1$. If $g_j = 0$ for $j \neq j_h$, $(h = 1, \cdots, p-m)$, then on $a_1 b_1$ there are continuous functions $D_{\alpha\beta}(x)$, $E_{h\beta}(x)$ such that

$$f_\alpha' = D_{\alpha\beta} f_\beta, \tag{4.6}$$

$$g_{j_h} = E_{h\beta} f_\beta. \tag{4.7}$$

Let $f_\alpha = f_{\alpha\gamma}(x)$, $(\gamma = 1, \cdots, m)$ be a fundamental set of solutions of (4.6) on $a_1 b_1$. Suppose that $g_j^*(x)$ is a set of functions such that $g_j^* = 0$ for $j \neq j_h$, and the set $g_{j_h}^*$ is linearly independent of the m sets of functions $g_{j_h} = E_{h\beta} f_{\beta\gamma}$ on $a_1 b_1$. It follows that if $z_i = w_i$ is the solution of

$$\mathcal{M}_i[z] = -g_j^* B_{ji}, \qquad t_i[z] = 0, \tag{4.8}$$

then w_i is not of the form $T_{ji}\pi_{j\alpha} f_\alpha$ on $a_1 b_1$, and hence $\int_a^b w_i K_{ij} w_j dx > 0$. Now one may choose an infinity of sets $g_{js}^*(x)$, $(s = 1, \cdots)$, such that if r is a given integer and $(d_t) \neq (0_t)$ $(t = 1, \cdots, r)$, then the set $g_j^* = g_{jt}^*(x) d_t$ satisfies the conditions described above for the set g_j^*. If $z_i = w_{is}$ are the corresponding solutions of (4.8), we have that the set w_{is}, $(s = 1, \cdots)$, satisfies the conditions of Theorem 4.1, and hence (2.2) has infinitely many characteristic values.

We have, in particular, the following result:

COROLLARY. *Suppose that* (2.2) *satisfies* (H_1)–(H_5), *the functions* B_{ij} *are of class* C^1 *on* ab, *and* $m < n - m$. *Then this system has infinitely many characteristic values.*

To establish this corollary, one need only note that the rank of (4.3) is not less than the greater of the two values m and $n - m$.

5. **A special boundary problem.** Consider the differential equations

$$\begin{aligned} \eta_\sigma' &= \mathcal{A}_{\sigma\tau}(x)\eta_\tau + \mathcal{B}_{\sigma\tau}(x)\zeta_\tau, \qquad \sigma, \tau = 1, \cdots, N, \\ \zeta_\sigma' &= \mathcal{C}_{\sigma\tau}(x)\eta_\tau - \mathcal{A}_{\tau\sigma}(x)\zeta_\tau - \lambda K_{\sigma\tau}(x)\eta_\tau, \end{aligned} \tag{5.1}$$

where $\|\mathcal{B}_{\sigma\tau}\|$, $\|\mathcal{C}_{\sigma\tau}\|$, $\|K_{\sigma\tau}\|$ are symmetric. The coefficients in (5.1) are real-valued and continuous on ab, and $u_\sigma K_{\sigma\tau} u_\tau > 0$ on ab for $(u_\sigma) \neq (0_\sigma)$. Associated with (5.1) we consider boundary conditions

$$(5.2)\qquad s_\gamma[\eta,\zeta] \equiv a^1_{\gamma\tau}\eta_\tau(a) - b^1_{\gamma\tau}\zeta_\tau(a) + a^2_{\gamma\tau}\eta_\tau(b) + b^2_{\gamma\tau}\zeta_\tau(b) = 0, \qquad \gamma = 1, \cdots, 2N,$$

where these conditions are linearly independent, and where the matrix $\|a^1_{\gamma\tau}b^1_{\delta\tau}+a^2_{\gamma\tau}b^2_{\delta\tau}\|$, $(\gamma,\ \delta=1,\cdots,2N)$, is symmetric. We shall also suppose that there is no non-identically vanishing solution η_σ, ζ_σ of (5.1), (5.2) with $\eta_\sigma\equiv 0$, $(\sigma=1,\cdots,N)$, on ab.

The system (5.1), (5.2) is self-adjoint with respect to the transformation matrix

$$\|T\| = \begin{Vmatrix} 0_{\sigma\tau} & \delta_{\sigma\tau} \\ -\delta_{\sigma\tau} & 0_{\sigma\tau} \end{Vmatrix}.$$

Under the above conditions,

$$\|B\| = \begin{Vmatrix} 0_{\sigma\tau} & 0_{\sigma\tau} \\ -K_{\sigma\tau} & 0_{\sigma\tau} \end{Vmatrix}, \quad \|S\| = \begin{Vmatrix} K_{\sigma\tau} & 0_{\sigma\tau} \\ 0_{\sigma\tau} & 0_{\sigma\tau} \end{Vmatrix}, \quad \|K\| = \begin{Vmatrix} 0_{\sigma\tau} & 0_{\sigma\tau} \\ 0_{\sigma\tau} & K_{\sigma\tau} \end{Vmatrix},$$

and (5.1), (5.2) satisfies (H_1)–(H_5) of §2. We may, therefore, assume without loss of generality that $\lambda=0$ is not a characteristic value, *and for simplicity we shall make this assumption.*

If we assume in addition that $\|\mathcal{B}_{\sigma\tau}\|$ has constant rank $N-q$, $(0\leqq q<N)$, on ab, this system is equivalent to a boundary problem associated with the second variation of a problem of Bolza in the calculus of variations which is normal and non-singular, but which does not necessarily satisfy the Clebsch condition. If $K_{\sigma\tau}=\delta_{\sigma\tau}$, we have the problem which Hu [5] considered, and for which he proved the existence of infinitely many characteristic values. His method of proof is closely related to that previously used by Bliss [2] in treating definitely self-adjoint problems. In the following, we shall not assume that $\|\mathcal{B}_{\sigma\tau}\|$ has constant rank $N-q$, *but simply that all the $\mathcal{B}_{\sigma\tau}$ are not identically zero on ab.*

The system adjoint to (5.1), (5.2) is

$$(5.3)\qquad \begin{aligned} u_\sigma' + u_\tau\mathcal{A}_{\tau\sigma} + v_\tau\mathcal{C}_{\tau\sigma} &= \lambda v_\tau K_{\tau\sigma}, \\ v_\sigma' + u_\tau\mathcal{B}_{\tau\sigma} - v_\tau\mathcal{A}_{\sigma\tau} &= 0, \end{aligned}$$

$$(5.4)\qquad b^1_{\gamma\tau}u_\tau(a) + a^1_{\gamma\tau}v_\tau(a) - b^2_{\gamma\tau}u_\tau(b) + a^2_{\gamma\tau}v_\tau(b) = 0, \qquad \gamma = 1,\cdots,2N.$$

The condition (3.4) then becomes, in terms of $(z_i)=(u_\sigma, v_\sigma)$,

$$(5.5)\qquad v_\sigma' + u_\tau\mathcal{B}_{\tau\sigma} - v_\tau\mathcal{A}_{\sigma\tau} = 0, \qquad \sigma = 1,\cdots,N,$$

and the integral of (3.7) reduces to

$$(5.6)\qquad \int_a^b v_\sigma K_{\sigma\tau}v_\tau dx.$$

Let a_1b_1 be a subinterval of ab on which all the $\mathcal{B}_{\sigma\tau}$ are not identically zero, and by $u_\tau = u_{\tau\nu}$, $(\nu = 1, \cdots, N+1)$, sets of functions of class C^1 on a_1b_1, with $u_{\tau\nu}(a_1) = 0 = u_{\tau\nu}(b_1)$, and such that the $N+1$ sets $u_{\tau\nu}\mathcal{B}_{\tau\sigma}$, $(\nu = 1, \cdots, N+1)$, are linearly independent on this interval. Let $v_\tau = v_{\tau\nu}(x)$ be the solution of (5.5) for $u_\tau = u_{\tau\nu}$, and satisfying the initial condition $v_{\tau\nu}(a_1) = 0$. Clearly the sets $v_\tau = v_{\tau\nu}(x)$, $(\nu = 1, \cdots, N+1)$, are linearly independent on a_1b_1. Now define $v_\tau = v_{\tau 1}(x)d_1 + \cdots + v_{\tau N+1}(x)d_{N+1}$, where the constants d_t, $(t = 1, \cdots, N+1)$, are not all zero and such that $v_\tau(b_2) = 0$. If the functions u_τ, v_τ thus defined on a_1b_1 are extended to the whole of ab by defining them as zero on the remainder of ab, then $(z_i) = (u_\tau, v_\tau)$ is an admissible arc for the above defined problem, and the corresponding integral (5.6) is positive.

Now let Δ_s, $(s = 1, 2, \cdots)$, be a denumerable set of non-overlapping subintervals of a_1b_1, and relative to Δ_s define a set of functions $(z_i) \equiv (w_{is}) \equiv (u_{\tau s}, v_{\tau s})$ in the manner described above. This set (w_{is}) is seen to satisfy the conditions of Theorem 4.1; hence the system (5.1) has a denumerable infinity of characteristic values.

6. **Oscillation and comparison theorems.** The following comparison theorem is a consequence of Theorem 5.5 of Reid [11].

THEOREM 6.1. *Suppose that the two boundary value problems*

$$\mathcal{L}_i[y] = \lambda B_{ij}y_j, \qquad s_i[y] \equiv M_{ij}y_j(a) + N_{ij}y_j(b) = 0, \tag{6.1}$$

$$\mathcal{L}_i[y] = \lambda B_{ij}y_j, \qquad s_i^*[y] \equiv M_{ij}^*y_j(a) + N_{ij}^*y_j(b) = 0, \tag{6.2}$$

are self-adjoint with respect to the same transformation matrix $\|T_{ij}\|$, *that each of these problems satisfies* (H_1)–(H_5), *and* $\lambda = 0$ *is not a characteristic value for either of these problems. For arbitrary values* $L > 0$, *let* V_L $[W_L]$ *denote the number of characteristic values* λ *of* (6.1) *such that* $|\lambda| < L$ $[|\lambda| \leqq L]$. *The numbers* V_L^*, W_L^* *are defined for* (6.2) *in an analogous manner. If the matrix*

$$\left\| \begin{matrix} M_{ij} & N_{ij} \\ M_{ij}^* & N_{ij}^* \end{matrix} \right\|$$

has rank $n+h$, *then for arbitrary* $L > 0$ *we have* $|V_L - V_L^*| \leqq h$, $|W_L - W_L^*| \leqq h$.

We shall say that a point x' on $a < x' \leqq b$ is a conjugate point of $x = a$ relative to the differential equations

$$\mathcal{L}_i[y] = \lambda B_{ij}(x)y_j \tag{6.3}$$

for a value λ, if, using the definition of Reid [11], §6, the point x' is a conjugate point of $x = a$ relative to the differential equations of (3.12) for the value $\Lambda = \lambda^2$. It is to be remarked that the hypotheses of §2 do not in general

imply that the order of abnormality of the differential equations of (3.12) on subintervals ab' is constant for $a<b'\leqq b$. Hence, as pointed out in Reid [11], in order to obtain the usual oscillation theorem it is necessary to use the extended definition of conjugate point there introduced. As a result of Theorem 3.4 above, and the oscillation and comparison theorems for the system (3.12) (see Reid [11], §§5, 6, and 7), we have the following result:

THEOREM 6.2. *Suppose that the system* (6.1) *satisfies hypotheses* (H_1)–(H_6) *and that* $\lambda=0$ *is not a characteristic value of this problem. Then for arbitrary values* $\lambda\neq 0$ *there are on the interval* $a<x<b$ *at least* $V_{|\lambda|}-n$ *and at most* $V_{|\lambda|}$ *conjugate points of* $x=a$ *relative to* (6.3) *for the given* λ.

In case the following additional assumption is satisfied, the conjugate points are given by the zeros of a certain determinant.

(H_7) If b' is an arbitrary value on $a<b'\leqq b$, there is no solution of $\mathcal{L}_i[y]=0$ satisfying $B_{ij}(x)y_j(x)\equiv 0$ on ab' except the identically vanishing solution $y_i\equiv 0$.

The statement (H_7) is the phrasing in terms of the system (6.3) of the condition that the differential equations of (3.12) are normal on every subinterval ab'. In this case the conjugate points of $x=a$ relative to (3.12) for a value Λ are the zeros of $|z_{ij}(x)|$, where

$$z_i=z_{ij}(x),\ \zeta_i=\zeta_{ij}(x), \qquad j=1,\cdots,n,$$

are linearly independent solutions of the differential equations of (3.12) such that

$$z_{ij}(a)=0, \qquad i,j=1,\cdots,n.$$

THEOREM 6.3. *Suppose that the system* (6.1) *satisfies* (H_7) *in addition to the hypotheses of Theorem* 6.2. *If* $y_i=y_{ij}(x:\lambda)$, $(j=1,\cdots,n)$, *denote the solutions of* (6.3) *for which* $y_{ij}(a:\lambda)=\delta_{ij}$, *then for arbitrary values* $\lambda\neq 0$ *the determinant*

$$|\,y_{ij}(x:\lambda)-y_{ij}(x:-\lambda)\,| \tag{6.4}$$

has on $a<x<b$ *at least* $V_{|\lambda|}-n$ *and at most* $V_{|\lambda|}$ *zeros, each zero counted a number of times equal to its index for the system of linear homogeneous equations whose coefficients are the rows of the determinant* (6.4).

Theorem 6.3 is a ready consequence of the fact that

$$z_{ij}\equiv(1/\lambda)T_{ki}(x)[y_{kj}(x:\lambda)-y_{kj}(x:-\lambda)],\qquad \zeta_{ij}\equiv y_{ij}(x:\lambda)+y_{ij}(x:-\lambda)$$

is a linearly independent set of solutions of the differential equations of (3.12) for $\Lambda=\lambda^2$ such that $z_{ij}(a)=0$, $(i,j=1,\cdots,n)$.

BIBLIOGRAPHY

1. Birkhoff and Hestenes, *Natural isoperimetric conditions in the calculus of variations*, Duke Mathematical Journal, vol. 1 (1935), pp. 198–286.

2. Bliss, *A boundary value problem for a system of ordinary linear differential equations of the first order*, these Transactions, vol. 28 (1926), pp. 561–584.

3. Bliss, *Definitely self-adjoint boundary value problems*, these Transactions, vol. 44 (1938), pp. 413–428.

4. Hölder, *Die Lichtensteinsche Method für die Entwicklung der zweiten Variation, angewandt auf das Problem von Lagrange*, Prace Matematyczno-Fizyczne, vol. 42 (1935), pp. 307–346.

5. Hu, *The problem of Bolza and its accessory boundary value problem*, *Contributions to the Calculus of Variations* 1931–1932, The University of Chicago Press, pp. 361–443.

6. Ince, *Ordinary Differential Equations*, London, 1927.

7. Morse, *Sufficient conditions in the problem of Lagrange with variable end conditions*, American Journal of Mathematics, vol. 53 (1931), pp. 517–546.

8. Morse, *The Calculus of Variations in the Large*, American Mathematical Society Colloquium Publications, vol. 18, 1934.

9. Reid, *A boundary value problem associated with the calculus of variations*, American Journal of Mathematics, vol. 54 (1932), pp. 769–790.

10. Reid, *Analogues of the Jacobi condition for the problem of Mayer in the calculus of variations*, Annals of Mathematics, (2), vol. 35 (1934), pp. 836–848.

11. Reid, *An integro-differential boundary value problem*, American Journal of Mathematics, vol. 60 (1938), pp. 257–292.

12. Wiggin, *A boundary value problem of the calculus of variations*, *Contributions to the Calculus of Variations* 1933–1937, The University of Chicago Press, pp. 245–275.

UNIVERSITY OF CHICAGO,
CHICAGO, ILL.

METRIC SPACES AND POSITIVE DEFINITE FUNCTIONS*

BY

I. J. SCHOENBERG

1. **Introduction.** Let E_m denote the m-dimensional euclidean space and generally E_m^p the pseudo-euclidean space of m real variables with the distance function

$$(|x_1 - x_1'|^p + \cdots + |x_m - x_m'|^p)^{1/p}, \qquad p > 0.$$

As $p \to \infty$ we get the space E_m^∞ with the distance function $\max_{i=1,\cdots,m} |x_i - x_i'|$. Let, furthermore, l^p stand for the space of real sequences with the series of pth powers of the absolute values convergent. Similarly let L^p denote the space of real measurable functions in the interval (0, 1) which are summable to the pth power, while C shall mean the space of real continuous functions in the same interval. In all these spaces a distance function is assumed to be defined as usual.† L^2 is equivalent to the real Hilbert space $\mathfrak{H}$. The spaces E_m^p, l^p, and L^p are metric only if $p \geqq 1$, but we shall consider them also for positive values of $p < 1$. Finally, if $\mathfrak{S}$ is a (not necessarily metric) space with the distance function $\overline{PP'}$, we shall denote by $\mathfrak{S}(\gamma)$ the new space which arises by changing the distance function from $\overline{PP'}$ to $\overline{PP'}^\gamma$, $(\gamma > 0)$.

A general theorem of Banach and Mazur ([1], p. 187) states that any separable metric space $\mathfrak{S}$ may be imbedded isometrically in the space C. Furthermore, as a special case of a well known theorem of Urysohn, any such space $\mathfrak{S}$ may be imbedded *topologically* in $\mathfrak{H}$. *Isometric* imbeddability of $\mathfrak{S}$ in $\mathfrak{H}$ is, however, a much more restricted property of $\mathfrak{S}$.

The chief purpose of this paper is to point out the intimate relationship between the problem of isometric imbedding and the concept of positive definite functions, if this concept is properly enlarged. As a first approach to this connection we consider here isometric imbedding in Hilbert space only. It turns out that the possibility of imbedding‡ in $\mathfrak{H}$ is very easily expressible in terms of the elementary function e^{-t^2} and the concept of positive definite functions (Theorem 1). The author's previous result ([10]) to the effect that $\mathfrak{H}(\gamma)$, $(0 < \gamma < 1)$, which is the space arising from $\mathfrak{H}$ by raising its metric to a

* Presented to the Society, December 29, 1937; received by the editors December 14, 1937.

† See, for example, Banach [1], pp. 11–12. The numbers in square brackets refer to the list of references at the end of the paper.

‡ Here and below the word "imbedding" stands for "isometrical imbedding."

fractional power, is imbeddable in $\mathfrak{H}$,* appears again as a simple consequence (Corollary 1). For the class of spaces $\mathfrak{S}_m$ arising from the euclidean space E_m by a general change of metric of the vector type (11) below, the condition of imbeddability in $\mathfrak{H}$ is directly expressible in a simple way in terms of the *usual* concept of positive definite functions as described by Mathias and Bochner (Theorem 2). The solution of this problem for $m=1$ (the problem of "screw lines" in $\mathfrak{H}$, von Neumann and Schoenberg [8]) allows us now to derive purely analytical results in the theory of positive definite functions with which it is equivalent. Two readily defined classes of positive definite functions are completely determined (Theorems 3 and 4). In particular two new proofs are given for the already known fact that the function $\exp[-|x|^p]$ is positive definite for values of p in the range $0<p\leqq 2$, and not positive definite if $p>2$† (§5). One proof is geometrical, the other proof covering the case $0<p\leqq 2$ is analytical and may be read independently of the rest of this paper.

All previous results now allow us to conclude that the spaces $E_m^p(\gamma)$, $l^p(\gamma)$, and $L^p(\gamma)$ are imbeddable in $\mathfrak{H}$ for values of γ in the range $0<\gamma\leqq p/2$, where p is restricted to the range $0<p\leqq 2$ (Theorem 5). For $p=2$ we regain our previous result concerning $\mathfrak{H}(\gamma)$. It is interesting to compare and combine this result about L^p with a theoem of Banach and Mazur ([1], p. 203) concerning the linear dimensions of L^2 and L^q. According to this theorem L^2 is imbeddable in L^q if $q>1$. Now as $L^p(\gamma)$, $(0<p\leqq 2, 0<\gamma\leqq p/2)$, is imbeddable in $\mathfrak{H}$ or L^2, it follows that $L^p(\gamma)$, $(0<p\leqq 2, 0<\gamma\leqq p/2)$, is imbeddable in any L^q if $q>1$.

Similar as yet unsolved problems concerning the case of $p>2$ are shown to be equivalent to further knowledge as to the positive definite character of certain special functions of m variables. One of these unsolved problems suggests a probably possible way of extending an interesting theorem of L. M. Blumenthal on metric sets of four points to such sets of any number of points.

The reader primarily interested in the geometrical results of this paper may omit §4 and §5 entirely by taking the known Corollary 3 for granted.

2. **Positive definite functions.** A real continuous function $f(x_1, x_2, \cdots, x_m)$ which is defined for all real values of its variables and is even, that is for which $f(-x_1, \cdots, -x_m)=f(x_1, \cdots, x_m)$, is said to be positive definite (p.d.) if

$$\sum_{i,k=1}^{n} f(x_1^{(i)}-x_1^{(k)}, \cdots, x_m^{(i)}-x_m^{(k)})\rho_i\rho_k \geqq 0 \tag{1}$$

* W. A. Wilson [12], had previously remarked that $E_1(1/2)$ is imbeddable in $\mathfrak{H}$.

† Due in various parts to Pólya, Mathias, and P. Lévy. For references see Bochner [3], pp. 76–77. For still another treatment of the case $0<p\leqq 2$ see Bochner [5].

for arbitrary real ρ_i and any n points $(x^{(i)})$, $(i=1, \cdots, n)$, for $n=2, 3, \cdots$. For $n=2$, (1) gives $|f(x_1, \cdots, x_m)| \leqq f(0, \cdots, 0)$, hence p.d. functions are bounded throughout space and take their maximum value at the origin. We shall use in the sequel only the following most simple properties of this interesting class of functions.

I. The function defined by

$$f(x_1, \cdots, x_m) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \psi(u_1, \cdots, u_m) \cdot \exp\left[i(x_1u_1 + \cdots + x_mu_m)\right] du_1 \cdots du_m, \tag{2}$$

where ψ is a non-negative even function such that its integral over the whole space exists, is p. d.

II. Any finite linear combination of p.d. functions with non-negative coefficients is again p.d. The product of two p.d. functions is again p.d.

III. A continuous function which is the limit of a sequence of p.d. functions is itself p.d.

For completeness we sketch the simple proofs. Property I follows from the fact that the left-hand side of (1) reduces to

$$\int \cdots \int \left| \sum_{k=1}^{n} \rho_k \exp\left[i(x_1^{(k)}u_1 + \cdots + x_m^{(k)}u_m)\right] \right|^2 \psi du_1 \cdots du_m \geqq 0.$$

The additive property is clear from the fact that (1) is a linear inequality in f. The multiplicative property is a direct consequence of a lemma of I. Schur ([11], p. 10)* which states that if $\sum_1^n a_{ik}\rho_i\rho_k$, $\sum_1^n b_{ik}\rho_i\rho_k$, are two positive quadratic forms, then $\sum_1^n a_{ik}b_{ik}\rho_i\rho_k$ is also positive. Property III is immediately clear by continuity. We shall later on use the fact that $f(x)=\cos \lambda x$ is a p.d. function.

A closely allied concept is as follows. Let $\mathfrak{S}$ be a space in which a distance function $\overline{PP'}$ is defined subject to the following conditions: (1) $\overline{PP'}=\overline{P'P}\geqq 0$ for arbitrary points P, P' in $\mathfrak{S}$, (2) $\overline{PP}=0$. A real continuous even function $g(t)$, which is defined in the range of values of $\pm\overline{PP'}$, (P, P' in $\mathfrak{S}$), is said to be positive definite in $\mathfrak{S}$ if

$$\sum_{i,k=1}^{n} g(\overline{P_iP_k})\rho_i\rho_k \geqq 0 \tag{3}$$

for arbitrary real ρ_i and any n points P_i (different or not) of $\mathfrak{S}$, $(n=2,3, \cdots)$.

* For a discussion and consequences of Schur's result see also Pólya and Szegö [9], pp. 106–107, 307–308.

This class of functions $g(t)$ has for a given space $\mathfrak{S}$ properties II and III above for similar reasons. Both definitions agree if $m=1$, while $\mathfrak{S}$ is the one-dimensional euclidean space E_1.

The peculiar relationship between the definitions, which will be clearer later on, is already exhibited by the following simple example needed in the sequel. From the formula

$$e^{-x^2} = 2^{-1}\pi^{-1/2}\int_{-\infty}^{\infty} e^{ixu}e^{-u^2/4}du$$

we get, replacing x by x_j, $(j=1, \cdots, m)$, and multiplying the resulting equations,

$$\exp\left[-(x_1^2 + \cdots + x_m^2)\right] = \exp\left[-\left[(x_1^2 + \cdots + x_m^2)^{1/2}\right]^2\right]$$

$$= 2^{-m}\pi^{-m/2}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\exp\left[i(x_1u_1 + \cdots + x_mu_m)\right] \tag{4}$$

$$\cdot \exp\left[-(u_1^2 + \cdots + u_m^2)/4\right]du_1 \cdots du_m,$$

which shows at a glance (property I) that the function

$$f = \exp\left[-(x_1^2 + \cdots + x_m^2)\right]$$

is p.d. and also that $g(t)=e^{-t^2}$ is p.d. in E_m. As m is arbitrary, this implies that *the function e^{-t^2} is positive definite in the real Hilbert space $\mathfrak{H}$.*

3. **Conditions for isometric imbedding in Hilbert space in terms of positive definite functions.** It was pointed out by K. Menger and by the author (for references see [10]) that a necessary and sufficient condition that a separable space $\mathfrak{S}$ be imbeddable in $\mathfrak{H}$ is that for any $n+1$ points of $\mathfrak{S}$, $(n\geqq 2)$, we have

$$\sum_{i,k=1}^{n}(\overline{P_0P_i}^2 + \overline{P_0P_k}^2 - \overline{P_iP_k}^2)\rho_i\rho_k \geqq 0,$$

for arbitrary real ρ_i.* Let us now put this condition in a slightly more symmetrical form. By summing over the three terms separately, we may write this as

$$2\sum_{1}^{n}\rho_k\cdot\sum_{1}^{n}\overline{P_0P_k}^2\rho_k - \sum_{1}^{n}\overline{P_iP_k}^2\rho_i\rho_k \geqq 0,$$

* This was proved for the case when $\mathfrak{S}$ is a separable *semi-metric* space; that is, when the metric $\overline{PP'}$ satisfies the additional condition (3) $\overline{PP'}>0$ if $P\neq P'$, whereas we postulated only that (1) $\overline{PP'}=\overline{P'P}\geqq 0$, (2) $\overline{PP}=0$. However, our quadratic inequality, for $n=2$, insures the triangle inequality $\overline{PQ}+\overline{QR}\geqq\overline{PR}$ for any three points of $\mathfrak{S}$. If we now *identify* with P all points Q such that $\overline{PQ}=0$ (which is now allowed, since $\overline{PQ}=0$ implies $\overline{RP}=\overline{RQ}$ for any R, on account of the triangle inequality) and do this for all points of $\mathfrak{S}$, we get a new space which is not only semi-metric but even metric.

and if we set $\rho_0 = -\sum_1^n \rho_k$, this last inequality is equivalent to

$$-\rho_0^2\overline{P_0P_0}^2 - 2\sum_0^n \overline{P_0P_k}^2\rho_0\rho_k - \sum_1^n \overline{P_iP_k}^2\rho_i\rho_k \geqq 0,$$

or, finally,

$$\sum_{i,k=0}^{n} \overline{P_iP_k}^2\rho_i\rho_k \leqq 0. \tag{5}$$

Hence the inequality (5), as a consequence of the relation

$$\sum_{i=0}^{n} \rho_i = 0, \tag{6}$$

is equivalent to the above stated condition of imbeddability.

Now we are prepared to express this condition in terms of p.d. functions. We have seen in the previous section that the function e^{-t^2} is p.d. in $\mathfrak{H}$. Clearly also $e^{-\lambda^2 t^2}$ (λ real) is p.d. in $\mathfrak{H}$ as can be seen from (4) or by direct reasoning. Hence $e^{-\lambda^2 t^2}$ is also p.d. in any subset of $\mathfrak{H}$. Now if $\mathfrak{S}$ is to be imbeddable in $\mathfrak{H}$, it is clearly a necessary condition that $e^{-\lambda^2 t^2}$ be p.d. in $\mathfrak{S}$.

Let us prove now that this condition, together with the separability of $\mathfrak{S}$, is also sufficient. To prove this we have to show that (5) holds as a consequence of (6) for any $n+1$ points P_i of $\mathfrak{S}$. Now as $e^{-\lambda^2 t^2}$ is p.d. in $\mathfrak{S}$, we have

$$\sum_0^n \rho_i\rho_k \exp\left[-\lambda^2\overline{P_iP_k}^2\right] \geqq 0 \tag{7}$$

by (3). We complete the proof in two different ways.

First proof. By expanding the left-hand side of (7) in power series we have, in view of (6),

$$-\lambda^2\left(\sum_0^n \overline{P_iP_k}^2\rho_i\rho_k\right) + \lambda^4\left(\sum_0^n \overline{P_iP_k}^4\rho_i\rho_k\right)/2 - \cdots \geqq 0,$$

which clearly implies (5) for small values of λ.

Second proof. Using the formulas

$$t^\alpha = c(\alpha)\int_0^\infty (1 - e^{-\lambda^2 t^2})\lambda^{-1-\alpha}d\lambda,$$

$$c(\alpha) = \left[\int_0^\infty (1 - e^{-\lambda^2})\lambda^{-1-\alpha}d\lambda\right]^{-1}, \qquad 0 < \alpha < 2,\ t \geqq 0, \tag{8}$$

which are immediately proved by substituting in the first integral λt^{-1} for λ, we have

$$\overline{P_iP_k}^{\alpha} = c(\alpha)\int_0^{\infty}(1 - \exp\,[-\lambda^2\overline{P_iP_k}^2])\lambda^{-1-\alpha}d\lambda.$$

If now ρ_i are any real numbers satisfying (6), we get, for $0<\alpha<2$,

$$(9)\qquad \sum_{i,k=0}^{n}\overline{P_iP_k}^{\alpha}\rho_i\rho_k = -c(\alpha)\int_0^{\infty}\left\{\sum_0^n \rho_i\rho_k \exp\,[-\lambda^2\overline{P_iP_k}^2]\right\}\lambda^{-1-\alpha}d\lambda \leqq 0,$$

by (7) and the obvious fact that $c(\alpha)>0$. Now we get again the desired inequality (5) on allowing α in (9) to approach the limit 2. We have thus proved the following theorem:

THEOREM 1. *A necessary and sufficient condition that a separable space $\mathfrak{S}$ with a distance function $\overline{PP'}$ with the properties $\overline{PP'}=\overline{P'P}\geqq 0$, $\overline{PP}=0$, be isometrically imbeddable in $\mathfrak{H}$, is that the family of functions $e^{-\lambda t^2}$, $(\lambda>0)$, be positive definite in $\mathfrak{S}$.*

Notice that the condition of this theorem may be restricted to require that $e^{-\lambda t^2}$ be p.d. in $\mathfrak{S}$ only for a set of positive values of λ admitting the origin $\lambda=0$ as a point of accumulation. The properties II and III (§2) will then imply that $e^{-\lambda t^2}$ is p.d. in $\mathfrak{S}$ for all positive values of λ.

Recalling that we denote by $\mathfrak{S}(\gamma)$, $(\gamma>0)$, the space obtained from $\mathfrak{S}$ by replacing its metric $\overline{PP'}$ by $\overline{PP'}^{\gamma}$, it is of interest to point out the further fact, implicitly contained in the previous second proof, that if $\mathfrak{S}$ is imbeddable in $\mathfrak{H}$, then so is $\mathfrak{S}(\gamma)$ for any value of γ in the range $0<\gamma<1$. Indeed, let $\alpha=2\gamma$; if $\mathfrak{S}$ is imbeddable in $\mathfrak{H}$, then $e^{-\lambda t^2}$, $(\lambda>0)$, is p. d. in $\mathfrak{S}$; hence (9) holds in virtue of (6), and $\mathfrak{S}(\gamma)$ is therefore imbeddable in $\mathfrak{H}$ on account of the form (5) of the imbeddability condition. Applying this conclusion to $\mathfrak{S}=\mathfrak{H}$ itself, we have the following corollary:

COROLLARY 1. *The space $\mathfrak{H}(\gamma)$, $(0<\gamma<1)$, obtained from Hilbert space $\mathfrak{H}$ by raising its metric to a power γ, is imbeddable in $\mathfrak{H}$.**

* We may even state the following more general theorem: *Let*

$$(8')\qquad F(t) = \int_0^{\infty}(1 - e^{-\lambda^2 t^2})\lambda^{-2}d\sigma(\lambda),$$

where $\sigma(\lambda)$ is non-decreasing for $\lambda\geqq 0$ and is such that $\int_1^{\infty}\lambda^{-2}d\sigma(\lambda)$ exists. If we change the metric of $\mathfrak{H}$ from $\overline{PP'}$ to $[F(\overline{PP'})]^{1/2}$, then the new space thus arising is imbeddable in $\mathfrak{H}$. Indeed (6), (7), and (8') imply

$$\sum_1^n F(\overline{P_iP_k})\rho_i\rho_k = -\int_0^{\infty}\left(\sum_{i,k=1}^{n}\rho_i\rho_k \exp\,[-\lambda^2\overline{P_iP_k}^2]\right)\lambda^{-2}d\sigma(\lambda) \leqq 0,$$

and the theorem follows on account of the form (5) and (6) of the imbeddability condition. We leave open the question whether or not (8') gives the most general function $F(t)$ with this property.

Added in proof, August, 1938: Formula (8') gives indeed all functions with the property stated above. See the following paper *Metric spaces and completely monotone functions*, to appear in the Annals of Mathematics.

We now turn our attention to a special class of spaces of the type $\mathfrak{S}$ which are defined as follows. Let $\phi(x_1, \cdots, x_m)$ be a continuous function defined for all real values of its variables with the following properties

$$\phi(x_1, \cdots, x_m) \geqq 0, \qquad \phi(0, \cdots, 0) = 0, \tag{10}$$
$$\phi(-x_1, \cdots, -x_m) = \phi(x_1, \cdots, x_m);$$

and call $\mathfrak{S}_m$ the space of points $P=(x_1, \cdots, x_m)$ with the distance function

$$\overline{PP'} = [\phi(x_1 - x_1', \cdots, x_m - x_m')]^{1/2}. \tag{11}$$

For such a space, which is obviously separable, the condition of Theorem 1 that $e^{-\lambda t^2}$ be p.d. in $\mathfrak{S}_m$ may be expressed in a more familiar form. Indeed, on comparing the formulas (1), (3), and (11), we now obtain from Theorem 1 the following theorem:

THEOREM 2. *The space $\mathfrak{S}_m$ of m real numbers with the metric* (11) *is imbeddable in $\mathfrak{H}$ if and only if the functions*

$$f_\lambda = \exp[-\lambda\phi(x_1, \cdots, x_m)], \qquad \lambda > 0, \tag{12}$$

are positive definite in the sense (1) *of Mathias and Bochner for all positive values of* λ.

If $\phi(x_1, \cdots, x_m)$ is a homogeneous function, the conditions* (12) *may be replaced by the single condition that the function*

$$f_1 = \exp[-\phi(x_1, \cdots, x_m)] \tag{13}$$

be positive definite.

We mention the following corollary:

COROLLARY 2. *If $\phi(x_1, \cdots, x_m)$ is homogeneous and such that $e^{-\phi}$ is positive definite, then*

$$e^{-\phi^\gamma}, \qquad 0 < \gamma < 1,$$

is also positive definite.

For if $e^{-\phi}$ is p.d., then by Theorem 2, $\mathfrak{S}_m$ is imbeddable in $\mathfrak{H}$ and therefor $\mathfrak{S}_m(\gamma)$ is also (Corollary 1). Hence $\exp[-\phi^\gamma]$ is seen to be p.d. by applying Theorem 2 the other way around.

4. **Determination of certain classes of positive definite functions.** In this section we shall assume $m=1$, $\phi(x)$ being therefore a continuous non-negative even function vanishing at the origin. In this case we know precisely when $\mathfrak{S}_1$,

* We say that ϕ is homogeneous of degree κ if $\phi(tx_1, \cdots, tx_m) = t^\kappa\phi(x_1, \cdots, x_m)$ holds identically in the x_i and for $t>0$. A continuous homogeneous function ϕ with the properties (10) must, unless it vanishes identically, have a positive degree κ.

which is the space of real numbers with the metric $[\phi(x-x')]^{1/2}$ is imbeddable in $\mathfrak{H}$ (von Neumann and Schoenberg [3]). The most general function $\phi(x)$ for which this is the case is of the form

$$\phi(x) = \int_0^\infty \frac{\sin^2 xu}{u^2}\, d\sigma(u), \tag{14}$$

where $\sigma(u)$ is non-decreasing for $u \geqq 0$ such that

$$\int_1^\infty \frac{d\sigma(u)}{u^2} \text{ exists.} \tag{15}$$

Hence by Theorem 2 we infer that if

$$[f(x)]^\lambda = e^{-\lambda\phi(x)}$$

is p.d. for $\lambda > 0$, then $\phi(x)$ is of the form (14) and conversely. If now $f(x)$ is any positive function whose positive powers $[f(x)]^\lambda$ are all p.d., then considering

$$f(x)/f(0) = e^{-\phi(x)},$$

where $\phi(x)$ is necessarily non-negative, even, and vanishing at the origin, we conclude as before that $\phi(x)$ is of the form (14). We have thus proved the following theorem:

THEOREM 3. *The most general positive function $f(x)$ whose positive powers $[f(x)]^\lambda$, $(\lambda > 0)$, are all positive definite is of the form*

$$f(x) = \exp\left\{c - \int_0^\infty \frac{\sin^2 xu}{u^2}\, d\sigma(u)\right\}, \tag{16}$$

where $\sigma(u)$ is a non-decreasing function subject to the restriction (15), *while c is any real constant.*

A few remarks are called for regarding the condition of this theorem that $[f(x)]^\lambda$ be p.d. for $\lambda > 0$. In the first place, as remarked after the statement of Theorem 1, the range of λ in this condition may be restricted to a sequence of positive numbers tending to zero.

A second and more important remark is that Theorem 3 becomes false if we assume only that the positive function $f(x)$ is p.d., or in other words: formula (16) always represents a positive and p.d. function; it does not, however, represent all such functions but only those whose fractional powers are also p. d. To prove this statement it suffices to exhibit a p.d. function $f(x) > 0$ such that $[f(x)]^\lambda$, $(0 < \lambda < 1)$, are not all p.d. Such a function is

$$f_\epsilon(x) = \epsilon^2 + \cos^2 x,$$

if ϵ is sufficiently small, for it is easy to see that

$$[f_\epsilon(x)]^{1/2} = (\epsilon^2 + \cos^2 x)^{1/2}$$

is not p.d. for such values of ϵ. Indeed, if it were p.d. for arbitrarily small ϵ, it would follow that the limiting function

$$\lim_{\epsilon \to 0} [f_\epsilon(x)]^{1/2} = |\cos x|$$

were also p.d. (property III). But this is not the case as is seen from the fact that the cosine series

$$|\cos x| = \frac{2}{\pi} + \frac{4}{\pi} \sum_{n=1}^{\infty} (-1)^{n-1} \frac{\cos 2nx}{4n^2 - 1}$$

has some negative coefficients, or even more directly from the fact that the form

$$\sum_{i,k=1}^{4} |\cos (x_i - x_k)| \rho_i \rho_k,$$

where $x_1 = 0$, $x_2 = \pi/4$, $x_3 = \pi/2$, $x_4 = 3\pi/4$, is not positive. Indeed, its determinant is readily found to be -1, hence negative. The essence of the matter is that Schur's theorem, "If $\sum a_{ik}\rho_i\rho_\kappa$ is positive, then so is $\sum a_{ik}^n \rho_i \rho_k$ for $n = 1, 2, 3, \cdots$," can not be so extended that we may conclude that $\sum |a_{ik}|^\lambda \rho_i \rho_k$, $(\lambda > 0)$, is positive.

The third and last remark is that Theorem 3 is now equivalent to the theorem that (14) and (15) give the most general $\phi(x)$ such that the space of real numbers with the metric $[\phi(x - x')]^{1/2}$ be imbeddable in $\mathfrak{H}$ (see [8]). Hence a direct proof of Theorem 3 would furnish a new proof of that theorem.

While the problem of determining all positive and positive definite functions is yet unsolved, there is another subclass of this class of functions which can now be readily determined. It will in fact be a subclass of the class determined by Theorem 3.

Let $\psi(x)$ be a p.d. function, c a real constant. Clearly

$$(17) \qquad f(x) = \exp [c + \psi(x)] = e^c \left(1 + \frac{1}{1!}\psi(x) + \frac{1}{2!}\psi^2(x) + \cdots\right)$$

is also p.d. in virtue of the properties II and III. It has moreover the following additional two properties: (α) It is bounded away from zero, since $f(x) \geqq \exp [c - \psi(0)] > 0$. (β) All its positive powers $[f(x)]^\lambda$, $(\lambda > 0)$, are also p.d. Let us show that the converse is true, that any function $f(x)$ having the properties (α) and (β) is of the form (17) where $\psi(x)$ is p.d. Now an $f(x)$ with the properties (α) and (β) clearly belongs to the class described by Theo-

rem 3, and all we have to do is to decide which functions of the form (16) are bounded away from zero. Now this is the case if and only if the exponent (14), namely

$$\phi(x) = \int_0^\infty \frac{\sin^2 xu}{u^2}\, d\sigma(u)$$

is bounded in $-\infty < x < \infty$. The characteristic conditions for this have likewise been determined by von Neumann and the author in [8] and they are as follows: $\phi(x)$ is bounded if and only if $\sigma(0) = \sigma(+0)$ and $\int_{+0}^{\infty} u^{-2} d\sigma(u)$ exists. But if these conditions are fulfilled, then the non-decreasing function

$$\tau(u) = \int_{+0}^{u} \frac{d\sigma(u)}{u^2}$$

is bounded for $u > 0$, and (14) can be written as

$$\begin{aligned} \phi(x) &= \int_{+0}^{\infty} \frac{\sin^2 xu}{u^2}\, d\sigma(u) = \int_{+0}^{\infty} \sin^2 xu\, d\tau(u) \\ &= \frac{1}{2}\int_{+0}^{\infty} (1 - \cos 2xu) d\tau(u) = \psi(0) - \psi(x), \end{aligned} \tag{18}$$

where

$$\psi(x) = \frac{1}{2}\int_{+0}^{\infty} \cos 2ux\, d\tau(u)$$

is p.d. Hence, by (14), (18), and (16),

$$f(x) = \exp\,[c - \phi(x)] = \exp\,[c - \psi(0) + \psi(x)]$$

and is indeed of the desired form (17). We have thus proved the following theorem:

THEOREM 4. *The most general positive function $f(x)$ which is bounded away from zero and whose positive powers $[f(x)]^\lambda$, $(\lambda > 0)$, are positive definite is of the form*

$$f(x) = \exp\,[c + \psi(x)],$$

where $\psi(x)$ is positive definite and c is a real constant.

EQUIVALENT STATEMENT: *If $f(x)$ is a positive function, then* $\log f(x)$ *differs by a constant from a positive definite function if and only if $f(x)$ is bounded away from zero and its powers $[f(x)]^\lambda$, $(\lambda > 0)$, are all positive definite.*

5. **On the positive definite character of** $\exp\,[-|x|^\nu]$. We shall need in the next section the following well known result:

COROLLARY 3. *The function*

$$\exp\,[-|x|^p] \tag{19}$$

is positive definite if $0<p\leqq 2$ *and not positive definite if* $p>2$.

First proof. We know that $\exp\,[-x^2]$ is p.d. from the formula just preceding (4) and property I. Hence $\exp\,[-|x|^{2\gamma}]$, $(0<\gamma<1)$, is p.d. by Corollary 2, as $\phi(x)=|x|^2$ is homogeneous, and the first part of the statement is proved. Suppose now that (19) were p.d. for a value $p>2$. By Theorem 2 this would imply that the space $E_1(p/2)$, which is the real axis with the metric $|x-x'|^{p/2}$, were imbeddable in $\mathfrak{H}$. But this is clearly impossible as $E_1(p/2)$ is not even a metric space since the distances among the points $x=0, 1, 2$ are $\overline{01}=1$, $\overline{12}=1$, $\overline{02}=2^{p/2}>2$, and they violate the triangle inequality.

Second direct proof for the case $0<p\leqq 2$. First we shall prove directly on the basis of properties II and III (§2) that the function $f(x)$ defined by (16) and (15) is p.d. By property III $(T\to\infty)$ it suffices to show that

$$\exp\left[-\int_0^T \frac{\sin^2 xu}{u^2}\,d\sigma(u)\right]$$

is p.d. By the definition of the Stieltjes integral, this is a limit of functions of the form

$$\exp\left[-\sum_{\nu=1}^{n} 2A_\nu \sin^2 xu_\nu\right], \qquad A_\nu>0,$$

and it suffices to show that each of the factors is p.d. Now

$$\exp\,[-2A\sin^2 xu]=\exp\,[-A]\,\exp\,[A\cos 2xu],$$

and this is p.d. because $\cos 2xu$ is p.d. and the exponential series has positive coefficients only. This point being disposed of, it suffices to show that $|x|^p$, $(0<p\leqq 2)$, is of the form (14).* This is apparent on account of the formula

$$|x|^p=\int_0^\infty \sin^2 xu\cdot u^{-1-p}du \Big/ \int_0^\infty \sin^2 u\cdot u^{-1-p}du,$$

valid for $0<p<2$, which is as easily established as the similar representation (8). An obvious step-function for $\sigma(u)$ in (14), with only one jump at the origin, settles the case $p=2$.

6. **On the imbedding of the spaces** E_m^p, l^p, L^p, $(0<p\leqq 2)$, **in** $\mathfrak{H}$ **by a change of metric.** We learned in §2 that e^{-t^2} is p.d. in $\mathfrak{H}$ or L^2. We shall now show

* It is not by accident that the function (19) is of the form (16) if $0<p\leqq 2$. For *if* (19) is p.d., its positive powers must also be p.d., and the function is necessarily of the form (16), by Theorem 3. Notice that it is not bounded away from zero and hence is not of the form (17).

that $e^{-|t|^p}$ *is p.d. in the spaces* E_m^p, l^p, *and* L^p *for values of* p *in the range* $0<p\leqq 2$.

The function $\exp[-|x|^p]$ was found to be p.d. if $0<p\leqq 2$. Hence the functions $\exp[-|x_i|^p]$, $(i=1, 2, \cdots, m)$, are also p.d. when regarded as functions of the m variables x_i, and we may infer, by property II (§2), that their product

$$f = \exp[-(|x_1|^p + \cdots + |x_m|^p)], \qquad 0 < p \leqq 2,$$

is p.d. But this is equivalent to the statement that $\exp[-|t|^p]$ is p.d. in E_m^p. As the similar statement for l^p is proved in the same way as for L^p, only requiring less care, we shall limit ourselves to the consideration of L^p. Let

$$P_i = x_i(t),$$

where

$$\int_0^1 |x_i(t)|^p dt, \qquad i = 1, 2, \cdots, n,$$

exists, be n points of L^p. For

$$\overline{P_iP_k} = \left(\int_0^1 |x_i(t) - x_k(t)|^p dt\right)^{1/p}$$

we have to show that

$$\sum_{i,k=1}^{n} \rho_i\rho_k \exp[-\overline{P_iP_k}^p] \geqq 0 \tag{20}$$

for real ρ_i. There is no loss of generality in assuming that the functions $x_i(t)$ are continuous, as continuous functions are everywhere dense in L^p. For

$$x_{\nu i} = m^{-1/p} \cdot x_i(\nu/m), \quad \nu = 1, 2, \cdots, m;\ i = 1, 2, \cdots, n,$$

we have

$$\left(\sum_{\nu=1}^{m} |x_{\nu i} - x_{\nu k}|^p\right)^{1/p} = \left(\frac{1}{m}\sum_{\nu=1}^{m} |x_i(\nu/m) - x_k(\nu/m)|^p\right)^{1/p} \tag{21}$$

$$\to \overline{P_iP_k}, \qquad m \to \infty,$$

which proves the inequality (20), for (20) is already known to hold for all values of m if the $\overline{P_iP_k}$ of (20) are replaced by the quantities on the left-hand side of (21).

The fact that $\exp[-|t|^p]$, $(0<p\leqq 2)$, is p.d. in E_m^p, l^p, L^p implies that $\exp[-|t|^2]$, and therefore also $\exp[-\lambda|t|^2]$, $(\lambda>0)$, is p.d. in $E_m^p(p/2)$, $l^p(p/2)$, and $L^p(p/2)$. We have thus proved the following theorem:

THEOREM 5. *The spaces E_m^p, l^p, and L^p, $(0<p\leqq 2)$, become imbeddable in $\mathfrak{H}$ if we raise their metrics to the power $p/2$ or less. In other words, $E_m^p(\gamma)$, $l^p(\gamma)$, and $L^p(\gamma)$ are imbeddable in $\mathfrak{H}$ for values of γ in the range $0<\gamma\leqq p/2$.**

Note that for $p=2$ this result coincides with Corollary 1.

The fact that E_m^p, $(0<p<1)$, which is not a metric space, may be made metric by raising its metric to a suitable power was already known in a different terminology. Indeed, the following readily established "substitute" for Minkowski's inequality ([7], p. 32)

$$\sum_1^m |x_i+y_i|^p \leqq \sum_1^m |x_i|^p + \sum_1^m |y_i|^p, \qquad 0<p<1,$$

means precisely that $E_m^p(p)$, and therefore $E_m^p(\gamma)$, $(0<\gamma\leqq p)$, is a metric space. Now Theorem 5 states that if we restrict this range of γ to $0<\gamma\leqq p/2$, then not only can any *three* points of $E_m^p(\gamma)$ be imbedded isometrically in a euclidean space, but the same is true for any number of such points, since the whole space is imbeddable in Hilbert space.

7. **Some unsolved problems.** We shall devote this last section to a few unsolved analytical problems and to a brief discussion of their geometrical implications.

PROBLEM 1. *Let p be a real number exceeding 2, and let $m=2, 3, \cdots$. Do there exist positive exponents κ such that the function*

$$\exp\left[-\left(|x_1|^p+\cdots+|x_m|^p\right)^\kappa\right] \tag{22}$$

is positive definite?

If there are such positive exponents, then Corollary 2 implies the existence of a positive number κ_m with the properties: (22) is p.d. if $0<\kappa\leqq\kappa_m$ and is not p.d. if $\kappa>\kappa_m$. Furthermore Corollary 3 implies that $p\kappa_m\leqq 2$ or $\kappa_m\leqq 2/p$. Theorem 2 furnishes the following geometrical equivalent to Problem 1. The function (22) is p.d. if and only if $E_m^p(p\kappa/2)$, and therefore also $l^p(p\kappa/2)$ and $L^p(p\kappa/2)$, are imbeddable in $\mathfrak{H}$.

Particularly interesting is the next problem concerning the limiting case $p\to\infty$.

---

* The following is a different phrasing of the fact that $L^p(p/2)$ is imbeddable in L^2: *If $0<p<2$, then there is a functional $y(t)=U[x(t)]$, defined for all functions $x(t)$ in L^p, with values $y(t)$ in L^2 such that the relation*

$$\int_0^1 |x_1(t)-x_2(t)|^p dt = \int_0^1 |y_1(t)-y_2(t)|^2 dt$$

holds for two arbitrary $x_1(t)$, $x_2(t)$ of L^p. This functional $y=U[x]$ is necessarily continuous and univalent. A study of its further properties might prove to be of interest in the theory of L^p.

PROBLEM 2. *Do there exist positive exponents γ such that the function*

$$\exp\,[-\,[\max\,(|\,x_1\,|,\,\cdots,\,|\,x_m\,|\,)]^{2\gamma}] = \min\,(e^{-|x_1|^{2\gamma}},\,\cdots,\,e^{-|x_m|^{2\gamma}}) \tag{23}$$

is positive definite?

Again, if such exponents exist, there would exist a positive number γ_m with the properties: (23) is p.d. if $0<\gamma\leqq\gamma_m$ and is not p.d. if $\gamma>\gamma_m$. Moreover the function (23) is p.d. if and only if $E_m^\infty(\gamma)$ is imbeddable in $\mathfrak{H}$.

The particular interest of this second problem is due to the following lemma which is due essentially to Fréchet ([6], pp. 161–162). Let us call a *finite metric set* and denote by σ_m a metric space composed of exactly $m+1$ distinct points.

LEMMA. *Any finite metric set σ_m of $m+1$ points may be imbedded isometrically in E_m^∞.*

Proof. Let $P_0,\,P_1,\,\cdots,\,P_m$ denote the points of σ_m and $\overline{P_iP_k}$ their distances. Consider in E_m^∞ the following $m+1$ points:

$$Q_i = (x_{1i},\,x_{2i},\,\cdots,\,x_{mi}) \equiv (\overline{P_1P_i},\,\overline{P_2P_i},\,\cdots,\,\overline{P_mP_i}), \qquad i = 0,\,1,\,\cdots,\,m.$$

For their distance $\overline{Q_iQ_k}$ in E_m^∞ we have

$$\overline{Q_iQ_k} = \max_{j=1,\,\cdots,\,m} |\,x_{ji} - x_{jk}\,| = \max_{j=1,\,\cdots,\,m} |\,\overline{P_jP_i} - \overline{P_jP_k}\,| = \overline{P_iP_k},$$

for certainly $|\,\overline{P_jP_i} - \overline{P_jP_k}\,| \leqq \overline{P_iP_k}$, $(j=1,\,\cdots,\,m)$, on account of the triangle inequality in σ_m, while the equality sign holds if j is equal to whichever of the numbers i or k happens to be different from zero and hence within the range of j. If both $i=k=0$, the result $\overline{Q_0Q_0}=0$ was clear from the start.

On the basis of this lemma it is natural to classify finite metric sets according to their "dimension" as follows: *A set σ_n is said to be of dimension m (always less than or equal to n, on account of the lemma) if it is imbeddable in E_m^∞, but not in E_{m-1}^∞. If the last requirement is removed, the dimension of σ_n does not exceed m.*

If Problem 2 were solved in the affirmative, we would get the following statement: If σ_n is any finite metric set of dimension less than or equal to m, then $\sigma_n(\gamma)$ is imbeddable in E_n if $0<\gamma\leqq\gamma_m$, and γ_m is the best constant.

This would generalize in a certain sense the following theorem due to L. M. Blumenthal ([2], p. 402): *If σ_3 is a finite metric set composed of four points, then $\sigma_3(\gamma)$ is imbeddable in E_3 if $0<\gamma\leqq 1/2$, and $1/2$ is the best constant.*

In conclusion let us point out the following perhaps not trivial remark concerning Problem 2: *γ_2 exists and is greater than or equal to $1/2$.* Indeed, in view of the formula

$$e^{-|t|} = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{e^{itx}}{1+x^2}\,dx,$$

we get by the transformation of variables $u=(\xi+\eta)/2$, $v=(\xi-\eta)/2$ in the double integral

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\,[i(xu+yv)]\,\frac{dudv}{[1+(u+v)^2][1+(u-v)^2]}$$
$$= \frac{1}{2}\int_{-\infty}^{\infty} \exp\,[i\xi(x+y)/2]\,\frac{d\xi}{1+\xi^2}\cdot\int_{-\infty}^{\infty} \exp\,[i\eta(x-y)/2]\,\frac{d\eta}{1+\eta^2}$$
$$= \frac{\pi^2}{2}\exp\,[-\tfrac{1}{2}|x+y| - \tfrac{1}{2}|x-y|\,] = \frac{\pi^2}{2}\exp\,[-\max(|x|,|y|)],$$

due to the relation $2\max(|x|, |y|) = |x+y| + |x-y|$. This formula shows that the function (23) is p.d. if $m=2$, $\gamma=1/2$. Hence this much is proved:

If σ_n is any finite metric set of dimension not exceeding 2, then $\sigma_n(\gamma)$, $(0<\gamma\leqq 1/2)$, is imbeddable in E_n.

If γ_3 exists, then it is certainly less than or equal to $1/2$, by Blumenthal's theorem, and is probably even less than $1/2$. A certain special set σ_4 shows that if γ_4 exists, it must be less than 0.45.

Bibliography

1. S. Banach, *Théorie des Opérations Linéaires*, Warsaw, 1932.

2. L. M. Blumenthal, *New theorems and methods in determinant theory*, Duke Mathematical Journal, vol. 2 (1936), pp. 396–404.

3. S. Bochner, *Vorlesungen über Fouriersche Integrale*, Leipzig, 1932.

4. ———, *Monotone Funktionen, Stieltjessche Integrale und harmonische Analyse*, Mathematische Annalen, vol. 108 (1933), pp. 378–410.

5. ———, *Stable laws of probability and completely monotone functions*, Duke Mathematical Journal, vol. 3 (1937), pp. 726–728.

6. M. Fréchet, *Les dimensions d'un ensemble abstrait*, Mathematische Annalen, vol. 68 (1910), pp. 145–168.

7. G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*, Cambridge, 1934.

8. J. von Neumann and I. J. Schoenberg, *Fourier integrals and metric geometry*, to appear in the Transactions of the American Mathematical Society.

9. G. Pólya and G. Szegö, *Aufgaben und Lehrsätze aus der Analysis*, vol. 2, Berlin, 1925.

10. I. J. Schoenberg, *On certain metric spaces arising from euclidean spaces by a change of metric and their imbedding in Hilbert space*, Annals of Mathematics, (2), vol. 38 (1937), pp. 787–793.

11. I. Schur, *Bemerkungen zur Theorie der beschränkten Bilinearformen mit unendlich vielen Veränderlichen*, Journal für die reine und angewandte Mathematik, vol. 140 (1911), pp. 1–28.

12. W. A. Wilson, *On certain types of continuous transformations of metric spaces*, American Journal of Mathematics, vol. 57 (1935), pp. 62–68.

Colby College,
Waterville, Me.

ON THE TRANSITIVITY OF PERSPECTIVITY IN CONTINUOUS GEOMETRIES*

BY

ISRAEL HALPERIN†

Introduction. The class of finite dimensional projective geometries has been extended to include non-finite dimensional ones by J. von Neumann's remarkable discovery of continuous geometries.‡ In an axiomatic formulation of the geometry as an irreducible complemented modular lattice§ the finiteness of the dimensionality is guaranteed by a chain condition. Von Neumann drops this chain condition and, retaining explicitly only two of its weak consequences, namely, completeness of the geometry and a certain continuity of the lattice operations, succeeds in establishing the existence of an essentially unique real-valued dimension function which may have either a discrete bounded range (the classical finite dimensional projective geometries) or a continuous bounded range (the new continuous geometries). In every case it is understood that the dimension function $D(a)$ is to satisfy

$$D(a+b)+D(ab)=D(a)+D(b) \tag{1}$$

for all a, b.

It is clear that such a dimension function will be closely connected with perspectivities. For a, b are said to be perspective if there exists a c such that

$$a+c=b+c, \qquad ac=bc;$$

and for such a, b (1) implies

$$\begin{aligned} D(a)+D(c) &= D(a+c)+D(ac) \\ &= D(b+c)+D(bc)=D(b)+D(c) \end{aligned}$$

and hence, if $D(c)$ is finite, $D(a)=D(b)$. This motivates a definition of equidimensionality, namely, a and b are called equidimensional if and only if they are perspective. That this definition will lead to the desired dimension function (in an irreducible system) depends in an essential way on the funda-

* Presented to the Society, December 30, 1936; received by the editors September 14, 1937.

† Sterling Research Fellow.

‡ See J. von Neumann: (1) Proceedings of the National Academy of Sciences, vol. 22 (1936), pp. 92–100, 101–108; (2) *Lectures on Continuous Geometry*, planographed, Institute for Advanced Study, Princeton, N. J., 1935–1937; (3) *Continuous Geometry*, American Mathematical Society Colloquium Lectures, to appear in book form. (2) will be referred to as C.G. The writer wishes to express his thanks to Professor von Neumann for many discussions of his new geometries.

§ See G. Birkhoff, Annals of Mathematics, vol. 36 (1935), pp. 743–748.

mental theorem that a, b equidimensional and b, c equidimensional together imply a, c equidimensional; in other words, that the relation of perspectivity is transitive.

The transitivity of perspectivity has been established by von Neumann for reducible as well as irreducible systems* but partly by indirect methods which require the full force of the completeness and continuity axioms. Now while these axioms are indeed necessary for the existence of the dimension function (in irreducible systems), weaker ones will secure the transitivity of perspectivity (in reducible as well as irreducible systems), in fact, just those parts of von Neumann's axioms which involve at most countable sets of elements.†

The present paper is devoted chiefly to a proof of the transitivity of perspectivity which uses direct methods throughout and holds for all systems satisfying these weaker axioms. The paper is divided into six sections. The weakened set of axioms to be used is formulated in §1. We require parts of C.G., part I, usually in very specialized form, and for convenience these are collected (briefly) in §§2, 3, 4. The new material in the proof of the transitivity of perspectivity is contained in §5. The additivity and continuity properties of perspectivity are established in §6. The Lemma 5.1 in §5 may perhaps be not without some interest of its own.

1. **The partially ordered system.** We shall consider a system L of elements $a, b, c, \cdots, x, y, u, v, \cdots, A, B, \cdots$ which is partially ordered, that is, we shall assume that a relation $a \leqq b$ (written equivalently $b \geqq a$) holds for certain pairs of elements of L in such a way that

(i) $a \leqq b$, $b \leqq c$ together imply $a \leqq c$, and

(ii) $a \leqq b$, $b \leqq a$ are together equivalent to $a = b$.

The following axioms are postulated:

AXIOM I. COUNTABLE COMPLETENESS. *For every finite or countably infinite set‡ of elements $a_1, a_2, \cdots$ there exist the following elements:*

I_1. *a sum element a (written $\sum_n a_n$ or equivalently $a_1 + a_2 + \cdots$) such that for any x of L, $x \geqq a$ if and only if $x \geqq a_n$ for every n,*

I_2. *an intersection element a (written $\prod_n a_n$ or equivalently $a_1 a_2 \cdots$) such that for any x of L, $x \leqq a$ if and only if $x \leqq a_n$ for every n.*

---

* For the general case see C.G., part III, p. 22, Theorem 2.3; the special (irreducible) case is also a consequence of the theorems of C.G., part I (see C.G., part I, p. 49, corollary to Theorem 5.16).

† That the "countable" axioms are really weaker than the original axioms of von Neumann can be shown by a simple example which satisfies the "countable" axioms but which has no zero, and hence is not complete.

‡ All sets considered in this paper will be non-void. Thus Axiom I does not imply the existence of a zero or of a unit element.

AXIOM II. COUNTABLE CONTINUITY. *Let $a_1, a_2, \cdots$ be any countably infinite sequence, and let c be an arbitrary element of L. Then*

II_1. $(\sum_n a_n)c = \sum_p \{(\sum_{n=1}^{p} a_n)c\}$;

II_2. $(\prod_n a_n) + c = \prod_p \{(\prod_{n=1}^{p} a_n) + c\}$.

AXIOM III. MODULARITY. *For all a, b, c,*

$$(a + b)c = \{a + (a + c)b\}c,$$

or what is equivalent, $a \leqq c$ implies $(a+b)c = a+bc$.

AXIOM IV. COMPLEMENTATION. *For any three a, b, c such that $a \leqq b \leqq c$ there exists an element d such that $b+d=c$, $bd=a$.*

2. **Independent sets of elements.** We make the following definition:

DEFINITION 2.1. *A finite ($\geqq 2$) or countably infinite set of elements $a_1, a_2, \cdots$ is independent (written $(a_n, n=1, 2, \cdots) \perp$) if for every two mutually exclusive subsets $a_{i_1}, a_{i_2}, \cdots$ and $a_{j_1}, a_{j_2}, \cdots$*

$$\left(\sum_n a_{i_n}\right)\left(\sum_n a_{j_n}\right) = \prod_n a_n.$$

*The a_n are said to be independent over θ if all such $(\sum_n a_{i_n})(\sum_n a_{j_n})$ equal θ.**

LEMMA 2.1. *If the a_n are independent over θ, then $\theta = \prod_n a_n$ and $(a_n, n=1, 2, \cdots) \perp$.*

Proof. Since $\prod_n a_n = a_1(\prod_{n \neq 1} a_n)$, the lemma follows from Definition 2.1.

LEMMA 2.2. *If $a_1, a_2, \cdots$ are independent over θ, then every subset $a_{i_1}, a_{i_2}, \cdots$ is independent over θ.*

Proof. The lemma follows directly from Definition 2.1 and Lemma 2.1.

LEMMA 2.3. *If $a_1, a_2, \cdots$ are independent over θ and if $(a_{i_r}, r=1, \cdots)$ are mutually exclusive subsets for $i=1, 2, \cdots$, then $\sum_r a_{i_r}$, $i=1, 2, \cdots$, are independent over θ.*

Proof. The lemma follows immediately from Definition 2.1.

LEMMA 2.4. *If $\theta, a_1, a_2, \cdots$ are such that for every two finite and mutually exclusive subsets $a_{i_1}, \cdots, a_{i_p}$ and $a_{j_1}, \cdots, a_{j_q}$*

$$\left(\sum_{n=1}^{p} a_{i_n}\right)\left(\sum_{n=1}^{q} a_{j_n}\right) = \theta,$$

then the a_n are independent over θ.

* If θ is a zero element of L, that is, if $a \geqq \theta$ holds for every a in L, then our independence over θ is precisely the notion of independence as used in C.G., part I, chap. 2.

Proof. Let a_{i_n}, $(n=1, \cdots)$, and a_{j_n}, $(n=1), \cdots$, be any two mutually exclusive subsets of $a_1, a_2, \cdots$. Then

$$\Big(\sum_n a_{i_n}\Big)\Big(\sum_n a_{j_n}\Big) = \Big(\sum_n a_{i_n}\Big)\Big\{\sum_q\Big(\sum_{n=1}^{q} a_{j_n}\Big)\Big\}$$

$$= \sum_q\Big\{\Big(\sum_n a_{i_n}\Big)\Big(\sum_{n=1}^{q} a_{j_n}\Big)\Big\}$$

$$= \cdots = \sum_q\Big\{\sum_p\Big(\sum_{n=1}^{p} a_{i_n}\Big)\Big(\sum_{n=1}^{q} a_{j_n}\Big)\Big\}$$

$$= \sum_q\Big\{\sum_p (\theta)\Big\} = \theta;$$

and the lemma follows from Definition 2.1.

COROLLARY. *A countably infinite set of elements is independent over θ if and only if every finite $(\geqq 2)$ subset is independent over θ.*

Proof. The corollary follows immediately from Lemmas 2.2 and 2.4.

LEMMA 2.5. *Let $\theta, a_1, a_2, \cdots$ satisfy $a_n \geqq \theta$ for every n. Let $r_1, r_2, \cdots, r_m$ be distinct integers, and let S be any set of integers not containing r_m. If $a_{r_m}\sum_{n\neq r_m} a_n = \theta$, then*

$$\Big(\sum_{n=1}^{m} a_{r_n}\Big)\Big(\sum_{n\in S} a_n\Big) = \Big(\sum_{n=1}^{m-1} a_{r_n}\Big)\Big(\sum_{n\in S} a_n\Big).$$

Proof.

$$\Big(\sum_{n=1}^{m} a_{r_n}\Big)\Big(\sum_{n\in S} a_n\Big) = \Big(\sum_{n=1}^{m-1} a_{r_n} + a_{r_m}\Big)\Big(\sum_{n\neq r_m} a_n\Big)\Big(\sum_{n\in S} a_n\Big)$$

$$= \Big\{\Big(\sum_{n=1}^{m-1} a_{r_n}\Big) + a_{r_m}\Big(\sum_{n\neq r_m} a_n\Big)\Big\}\Big(\sum_{n\in S} a_n\Big)$$

$$= \Big(\sum_{n=1}^{m-1} a_{r_n} + \theta\Big)\Big(\sum_{n\in S} a_n\Big) = \Big(\sum_{n=1}^{m-1} a_{r_n}\Big)\Big(\sum_{n\in S} a_n\Big),$$

which proves the lemma.

LEMMA 2.6. *If $\theta, a_1, a_2, \cdots$ are such that $a_{n+1}(a_1 + \cdots + a_n) = \theta$ for every $n=1, 2, \cdots$, then the a_n are independent over θ.*

Proof. By Lemma 2.4 we need only show that

$$\Big(\sum_{n=1}^{p} a_{i_n}\Big)\Big(\sum_{n=1}^{q} a_{j_n}\Big) = \theta$$

for all finite p, q and different $i_1, \cdots, i_p; j_1, \cdots, j_q$; and this follows from a finite number of applications of Lemma 2.5.

COROLLARY. *If $\theta, c, a_1, a_2, \cdots$ are such that*

$$a_n(a_{n+1} + \cdots + a_{n+p} + c) = \theta$$

for all $n \geqq 1$, $p \geqq 1$, then the c, a_n are independent over θ.

Proof. By Lemma 2.6, $c, a_{n+p}, a_{n+p-1}, \cdots, a_n$ are independent over θ for all $n \geqq 1$, $p \geqq 0$. Therefore, by the corollary to Lemma 2.4, $c, a_1, a_2, \cdots$ are independent over θ.

LEMMA 2.7. *Let $a_1, a_2, \cdots$ be independent over θ. If $S_1, S_2, \cdots$ are arbitrary subsets of the integers $1, 2, \cdots$ and S is the set of the integers common to all S_n, then*

$$\prod_t \left(\sum_{n \in S_t} a_n \right) = \sum_{n \in S} a_n.$$

Proof. Let $T = (a_{r_1}, a_{r_2}, \cdots)$ be the set of the integers not in S. Then

$$\begin{aligned}
\prod_t \left(\sum_{n \in S_t} a_n \right) &= \left(\sum_{n \in S} a_n + \sum_{n \in T} a_n \right) \prod_t \left(\sum_{n \in S_t} a_n \right) \\
&= \sum_{n \in S} a_n + \left(\sum_{n \in T} a_n \right) \prod_t \left(\sum_{n \in S_t} a_n \right) \\
&= \sum_{n \in S} a_n + \sum_m \left\{ \left(\sum_{t=1}^{m} a_{r_t} \right) \prod_{st} \left(\sum_{n \in S_t} a_n \right) \right\} \\
&= \sum_{n \in S} a_n + \sum_m (\theta) = \sum_{n \in S} a_n
\end{aligned}$$

(by repeated use of Lemma 2.5) as required.

LEMMA 2.8. *If a_n, $(n = 1, 2, \cdots)$, are independent over θ and $\theta \leqq u_n \leqq a_n$ for $n = 1, \cdots, p$, and if $\theta \leqq v_n \leqq a_n$ for $n = 1, \cdots, q$, then*

$$\left(\sum_{n=1}^{p} u_n \right) \left(\sum_{n=1}^{q} v_n \right) = \sum_{n=1}^{\min(p,q)} (u_n v_n).$$

Proof.

$$\begin{aligned}
(u_1 + u_2)(v_1 + v_2) &= (u_1 + u_2)(u_1 + a_2)(a_1 v_1 + v_2) \\
&= (u_1 + u_2)\{v_2 + v_1 a_1 (u_1 + a_2)\} \\
&= (u_1 + u_2)\{v_2 + v_1(u_1 + a_1 a_2)\} = (u_1 + u_2)(v_2 + v_1 u_1) \\
&= \cdots = (u_1 + u_2)(u_1 v_1 + u_2 v_2) = u_1 v_1 + u_2 v_2.
\end{aligned}$$

Thus the lemma holds for $p = q = 2$. But if the lemma holds for all $p = q < m$,

then it holds for $p=q=m$ too. For

$$\left(\sum_{n=1}^{m} u_n\right)\left(\sum_{n=1}^{m} v_n\right) = \left(\sum_{n=1}^{m-1} u_n\right)\left(\sum_{n=1}^{m-1} v_n\right) + u_m v_m = \sum_{n=1}^{m} (u_n v_n),$$

since $\theta \leqq \sum_{n=1}^{m-1} u_n$, $\sum_{n=1}^{m-1} v_n \leqq \sum_{n=1}^{m-1} a_n$; $\theta \leqq u_m$, $v_m \leqq a_m$; and $\sum_{n=1}^{m-1} a_n$, a_m are independent over θ. Thus the lemma holds for all $p=q$. If $p \neq q$, say $p<q$, we can set $u_n=\theta$ for $p<n \leqq q$ and apply the result just proved for $p=q$.

LEMMA 2.9. *If $a_1, a_2, \cdots$ are independent over θ and $\theta \leqq u_n$, $v_n \leqq a_n$; for $n=1, 2, \cdots$, then*

$$\left(\sum_n u_n\right)\left(\sum_n v_n\right) = \sum_n (u_n v_n).$$

Proof.

$$\begin{aligned}\left(\sum_n u_n\right)\left(\sum_n v_n\right) &= \left\{\sum_p \left(\sum_{n=1}^{p} u_n\right)\right\}\left\{\sum_q \left(\sum_{n=1}^{q} v_n\right)\right\} \\ &= \sum_q \left\{\sum_p \left(\sum_{n=1}^{p} u_n\right)\left(\sum_{n=1}^{q} v_n\right)\right\} \\ &= \sum_q \left\{\sum_p \left(\sum_{n=1}^{\min(p,q)} (u_n v_n)\right)\right\} \\ &= \sum_n (u_n v_n)\end{aligned}$$

(by Lemma 2.8) as required.

LEMMA 2.10. *Let $a_1, a_2, \cdots$ be independent over θ. If $a_i \geqq a_{ij} \geqq \theta$ for all i, j, and if the elements a_{ij}, $j=1, \cdots$, are independent over θ (whenever there are at least two elements in the set) for every $i=1, 2, \cdots$, then the set of all a_{ij}, $(i, j=1, \cdots)$, is independent over θ.*

Proof. If $a_{i_r j_r}$, $(r=1, \cdots)$, and $a_{k_s l_s}$, $(s=1, \cdots)$, are mutually exclusive subsets of the a_{ij}, then

$$\begin{aligned}\left(\sum_r a_{i_r j_r}\right)\left(\sum_s a_{k_s l_s}\right) &= \left\{\sum_{i_r} \left(\sum_{j_r} a_{i_r j_r}\right)\right\}\left\{\sum_{k_s} \left(\sum_{l_s} a_{k_s l_s}\right)\right\} \\ &= \sum_n \left\{\left(\sum_{j_r} a_{n j_r}\right)\left(\sum_{l_s} a_{n l_s}\right)\right\} \\ &= \sum_n \{\theta\} = \theta\end{aligned}$$

(by Lemma 2.9), which proves the lemma.

LEMMA 2.11. *If $\theta \leqq a_0$ and if a_n, a_n' are defined for $n=1, 2, \cdots$ in such a way that*

$$a_{n-1} = a_n + a_n', \qquad a_n a_n' = \theta,$$

for $n=1, 2, \cdots$, then $(\prod_r a_r)$, a_n', $(n=1, 2, \cdots)$, are independent over θ and $a_0 = \sum_n a_n' + \prod_r a_r$.

Proof.

$$a_n'\left(a_{n+1}' + a_{n+2}' + \cdots + a_{n+p}' + \prod_r a_r\right)$$

$$= a_n' a_n\left(a_{n+1}' + a_{n+2}' + \cdots + a_{n+p}' + \prod_r a_r\right) = \theta$$

for all n, p. Hence $(\prod_r a_r)$, a_1', a_2', $\cdots$ are independent over θ by the corollary to Lemma 2.6. Furthermore

$$a_0 = a_1' + a_1 = a_1' + a_2' + a_2 = \cdots$$

$$= \sum_{n=1}^{r} a_n' + a_r = \sum_{n=1}^{\infty} a_n' + a_r,$$

for $r=1, 2, \cdots$. Hence

$$a_0 = \prod_r\left(\sum_{n=1}^{\infty} a_n' + a_r\right) = \sum_{n=1}^{\infty} a_n' + \prod_{r=1}^{\infty} a_r;$$

and the lemma is proved.

3. **Perspectivities and perspective mappings.** We make the following definition:

DEFINITION 3.1. *a, b are perspective (written $a \sim b$) if there exists an element c such that*

(i) $a+c=b+c$,

(ii) $ac=bc$.

Then c is called the axis of the perspectivity.

LEMMA 3.1. *If a, b are perspective and $\theta \leqq ab$, then there exists an element d such that*

(i) $a+d=b+d=a+b$, *and*

(ii) $ad=bd=\theta$.

Proof. Let c be an axis of perspectivity for a and b. Since $\theta \leqq ab \leqq \{c(a+b)+ab\}$, Axiom IV secures the existence of an element d such that

$$d + ab = c(a+b) + ab, \qquad dab = \theta.$$

For this d we have

$$a + d = a + c(a + b) + ab = (a + c)(a + b) = a + b.$$

Similarly $b+d=a+b$, and (i) holds.

$$\begin{aligned} ad &= ad\{c(a + b) + ab\} = d\{ab + ac(a + b)\} \\ &= dab \\ &= \theta, \end{aligned}$$

since $ac(a+b)=ac\leqq ab$. Similarly $bd=\theta$, and (ii) holds. Thus d satisfies the requirements of the lemma.

DEFINITION 3.2. *The sublattice of the x satisfying $x\leqq a$ is denoted by $L(a)$. If $a\leqq b$, the sublattice of the x satisfying $a\leqq x\leqq b$ is denoted by $L(a, b)$.**

LEMMA 3.2. *If a, b are perspective with axis c and $\theta=ac=bc$, then a $(1, 1)$ correspondence between the elements of $L(\theta, a)$ and those of $L(\theta, b)$ which preserves the relation $\leqq$ is defined by the inverse mappings*

$$(P) \quad a_1 \rightarrow b_1 = (a_1 + c)b$$
$$(Q) \quad b_1 \rightarrow a_1 = (b_1 + c)a.$$

Proof. If $\theta\leqq x\leqq a$ then under (P) $x\rightarrow(x+c)b$, and under (Q)

$$\begin{aligned} (x + c)b \rightarrow \{(x + c)b + c\}a &= (x + c)(b + c)a \\ &= \{c + x(b + c)\}a = ca + x(b + c) \\ &= x(a + c) = x. \end{aligned}$$

Hence (Q) is inverse to (P). Similarly (P) is inverse to (Q). It follows that the correspondence is $(1, 1)$. The invariance of the relation $\leqq$ is clear from the definition of (P) and (Q).

DEFINITION 3.3. *The mappings of Lemma 3.2 are called perspective mappings.*

LEMMA 3.3. *If a_1 corresponds to b_1 under a perspective mapping, then $a_1\sim b_1$.*

Proof. Suppose a_1 corresponds to b_1 under a perspective mapping of $L(\theta, a)$ on $L(\theta, b)$ with axis c. Then a_1 is perspective to b_1 with axis c, for

$$\begin{aligned} a_1 + c &= (b_1 + c)a + c = (b_1 + c)(a + c) \\ &= (b_1 + c)(b + c) = b_1 + c, \\ a_1c &= a(b_1 + c)c = ac = bc = b(a_1 + c)c = b_1c, \end{aligned}$$

and conditions (i) and (ii) therefore hold.

* The Axioms I, II, III, IV hold in $L(a)$ and in $L(a, b)$. $L(a)$ has a unit (greatest) element, namely a, and $L(a, b)$ has a unit element b and a zero (smallest) element a.

LEMMA 3.4. *If P_i is a perspective mapping of $L(\theta, a_i)$ on $L(\theta, b_i)$ for $i=1, \cdots, p$, where $a_{i+1}=b_i$ for $i=1, \cdots, p-1$, then the product mapping of the P_i is a $(1, 1)$ mapping of $L(\theta, a_1)$ on $L(\theta, b_p)$ which preserves the relation $\leqq$.*

Proof. The lemma follows immediately from Definition 3.3.

DEFINITION 3.4. *The mapping of Lemma 3.4 is called a projective mapping.*

4. **Transitivity of perspectivity in special cases.** We prove the following lemma:

LEMMA 4.1. *$a\sim b$, $b\sim c$, (a, b, c) $\perp$ together imply $a\sim c$.*

Proof. By Lemma 3.1 x, y exist such that

$$a + x = b + x = a + b, \qquad b + y = c + y = b + c,$$
$$ax = bx = \theta, \qquad by = cy = \theta,$$

where $\theta = abc$.

Then a is perspective to c with axis $d=(a+c)(x+y)$. For

$$a + d = a + (a + c)(x + y) = (a + c)(a + x + y)$$
$$= (a + c)(a + b + y) = (a + c)(a + b + c) = a + c.$$

Similarly $c+d=a+c$. Thus $a+d=c+d$, and (i) holds. Also

$$ad = a(a + c)(x + y) = a(x + y) = a(a + b)(x + y)$$
$$= a\{x + (a + b)y(b + c)\} = a\{x + \theta\} = \theta.$$

Similarly $cd=\theta$. Thus $ad=cd$, and (ii) holds.

LEMMA 4.2. *$a_n\sim b_n$, for $n=1, 2, \cdots$, and $(a_n+b_n, n=1, 2, \cdots)$ $\perp$ together imply $\sum_n a_n \sim \sum_n b_n$.*

Proof. By Lemma 3.1 we may assume the existence of elements x_n such that

$$a_n + x_n = b_n + x_n = a_n + b_n,$$
$$a_n x_n = b_n x_n = \theta,$$

where $\theta = \prod_n (a_n b_n)$. Then $\sum_n a_n$, $\sum_n b_n$ are perspective with axis $\sum_n x_n$, for the relations

$$\sum_n a_n + \sum_n x_n = \sum_n (a_n + x_n)$$
$$= \sum_n (b_n + x_n) = \sum_n b_n + \sum_n x_n,$$

give property (i); and

$$\Big(\sum_n a_n\Big)\Big(\sum_n x_n\Big) = \sum_n (a_n x_n)$$

$$= \sum_n (b_n x_n) = \Big(\sum_n b_n\Big)\Big(\sum_n x_n\Big)$$

(by Lemma 2.9) gives property (ii).

LEMMA 4.3. *If an infinite independent sequence of elements* $a_0, a_1, \cdots$ *satisfy* $a_n \sim a_{n+1}$, *for* $n = 0, 1, \cdots$, *then* $a_0 = \prod_n a_n$.

Proof. From Lemmas 2.2 and 4.1, $a_0 \sim a_n$ for all n. By Lemma 3.1 we may therefore assume the existence of elements x_n, $(n = 1, 2, \cdots)$, such that

$$a_0 + x_n = a_n + x_n = a_0 + a_n,$$
$$a_0 x_n = a_n x_n = \theta,$$

where $\theta = \prod_n a_n$. We deduce successively

$$a_0 \leqq a_n + x_n, \qquad n = 1, 2, \cdots,$$

$$a_0 \leqq \Big(\sum_{n=p}^{\infty} a_n\Big) + \Big(\sum_{n=1}^{\infty} x_n\Big), \qquad p = 1, 2, \cdots,$$

$$a_0 \leqq \prod_p \Big\{\Big(\sum_{n=p}^{\infty} a_n\Big) + \Big(\sum_{n=1}^{\infty} x_n\Big)\Big\} = \Big\{\prod_p \Big(\sum_{n=p}^{\infty} a_n\Big)\Big\} + \Big(\sum_{n=1}^{\infty} x_n\Big)$$

$$= \sum_{n=1}^{\infty} x_n$$

by Lemma 2.7. Hence

$$a_0 = a_0 \sum_{n=1}^{\infty} x_n = a_0 \Big(\sum_p \sum_{n=1}^{p} x_n\Big) = \sum_p \Big(a_0 \sum_{n=1}^{p} x_n\Big) = \theta,$$

if only $a_0 \sum_{n=1}^{p} x_n = \theta$ for $p = 1, 2, \cdots$. Now for any fixed p,

$$a_0 \sum_{n=1}^{p} x_n = a_0 \Big\{\sum_{n=1}^{p-1} x_n + \Big(a_0 + \sum_{n=1}^{p-1} x_n\Big) x_p\Big\},$$

and, since

$$x_p \Big(a_0 + \sum_{n=1}^{p-1} x_n\Big) = x_p (a_0 + a_p)\Big(a_0 + \sum_{n=1}^{p-1} a_n\Big)\Big(a_0 + \sum_{n=1}^{p-1} x_n\Big) = x_p a_0 = \theta,$$

therefore

$$a_0 \sum_{n=1}^{p} x_n = a_0 \sum_{n=1}^{p-1} x_n.$$

A finite number of such reductions gives

$$a_0 \sum_{n=1}^{p} x_n = a_0 x_1 = \theta,$$

as required, and the lemma is proved.

LEMMA 4.4. *$a \sim x$, $x \sim a_1$, $ax \leqq a_1 \leqq a$ together imply $a_1 = a$.*

Proof. Let $ax = a_1 x = \theta$. Since $\theta \leqq a_1 \leqq a$, Axiom IV secures the existence of an element a_1' such that

$$a_1 + a_1' = a, \qquad a_1 a_1' = \theta.$$

By Lemmas 3.1 and 3.2 there exist perspective mappings T_1 of $L(\theta, a)$ on $L(\theta, x)$ and T_2 of $L(\theta, x)$ on $L(\theta, a_1)$. Define by induction on n

$$x_n = T_1(a_n), \quad x_n' = T_1(a_n'), \quad a_{n+1} = T_2(x_n), \quad a_{n+1}' = T_2(x_n').$$

Then Lemma 2.11, the relation $ax = \theta$, and Lemma 2.10, give $(a_n', x_n', n = 1, 2, \cdots) \perp$. Lemma 2.2 then gives $(a_n', x_n', a_{n+1}') \perp$, and Lemmas 3.3 and 4.1 give $a_n' \sim a_{n+1}'$, for $n = 1, 2, \cdots$. Lemma 2.2 shows that $(a_n', n = 1, 2, \cdots) \perp$; hence by Lemma 4.3 $a_1' = \prod_n a_n' = \theta$. Thus $a_1 = a_1 + \theta = a_1 + a_1' = a$; and the lemma is proved.

DEFINITION 4.1. *If θ has been defined, we sometimes write*

$$\sum_n (\oplus\, x_n)$$

(or the equivalent $x_1 \oplus x_2 \oplus \cdots$) in place of $\sum_n x_n$ (or $x_1 + x_2 + \cdots$), provided the x_n are independent over θ. If $\theta \leqq u \leqq v$, then $[v - u]$ will denote an element (fixed) such that $u \oplus [v - u] = v$. (Such an element exists by Axiom IV.)

LEMMA 4.5. *$a \sim x$, $x \sim b$, $ab \leqq x$ together imply $a \sim b$.*

Proof. (α) Consider the special case where $ab = bx = ax$ and $x \leqq a + b$. Let $b_1 = b(a + x)$. Then b_1 is perspective to x with axis a, for

$$b_1 + a = b(a + x) + a = (b + a)(a + x) = x + a,$$

hence relation (i) holds; and

$$b_1 a = b(a + x)a = ba = xa,$$

hence (ii) holds.

Since $b \sim x$ and $b_1 x = b(a + x)x = bx$, $bx \leqq b_1 \leqq b$; and Lemma 4.4 gives $b_1 = b$. Hence $b \leqq a + x$. Similarly $a \leqq b + x$. It follows that a is perspective to b with axis x, for

$$a + x = a + x + b = b + x;$$

hence relation (i) holds, and $ax = bx$ (by the special hypotheses of (α)) hence relation (ii) holds. This proves Lemma 4.5 in the special case (α).

(β) Suppose now only that $ab = bx = ax$ (equal, say θ). Let T_1, T_2 be perspective mappings of $L(\theta, x)$ on $L(\theta, a)$ and on $L(\theta, b)$, respectively. Set $a_0 = a$, $x_0 = x$, $b_0 = b$, and define a_n, a_n', x_n, x_n', b_n, b_n', for $n = 1, 2, \cdots$, by induction on n as follows:

$$\begin{aligned} x_n &= x_{n-1}(a_{n-1} + b_{n-1}), & a_n &= T_1(x_n), & b_n &= T_2(x_n), \\ x_n' &= [x_{n-1} - x_n], & a_n' &= T_1(x_n'), & b_n' &= T_2(x_n'). \end{aligned}$$

If we set $\bar{a} = \prod_n a_n$, $\bar{x} = \prod_n x_n$, $\bar{b} = \prod_n b_n$, then Lemma 2.11, the relation $ab = \theta$, and Lemma 2.3 give $a = \bar{a} + \sum_n a_n'$, $b = \bar{b} + \sum_n b_n'$; and $\bar{a} + \bar{b}$, $a_n' + b_n'$, $(n = 1, 2, \cdots)$, are independent over θ. Lemma 3.3 shows that $a_n' \sim x_n'$ and $x_n' \sim b_n'$. Since $a_n' b_n' = \theta$ and $x_n'(a_n' + b_n') = \theta$, Lemmas 2.6 and 4.1 show that $a_n' \sim b_n'$, for $n = 1, 2, \cdots$. Now Lemma 4.2 shows that $a \sim b$ if only $\bar{a} \sim \bar{b}$.

$\bar{a}$, $\bar{x}$, $\bar{b}$ satisfy the hypotheses of Lemma 4.5 and the special conditions of (α), for $\bar{a} = T_1(\bar{x})$, $\bar{b} = T_2(\bar{x})$ imply $\bar{a} \sim \bar{x}$, $\bar{x} \sim \bar{b}$; $\bar{a}\bar{x} = \bar{a}ax\bar{x} = \theta$, $\bar{b}\bar{x} = \bar{b}bx\bar{x} = \theta$, $\bar{a}\bar{b} = \bar{a}ab\bar{b} = \theta$; and, since $x_n \leqq a_{n-1} + b_{n-1}$ for all n,

$$\begin{aligned} \bar{x} &\leqq \prod_n (a_n + b_n) = \prod_q \prod_p \Big(\prod_{n=1}^{p} a_n + \prod_{n=1}^{q} b_n \Big) \\ &= \prod_n a_n + \prod_n b_n = \bar{a} + \bar{b} \end{aligned}$$

by two applications of Axiom II_2. Hence $\bar{a} \sim \bar{b}$; and Lemma 4.5 is proved for the special case (β).

(γ) Suppose now only $ab = bx$ (equal, say θ). Let T_1, T_2 be perspective mappings of $L(\theta, a)$ on $L(\theta, x)$ and of $L(\theta, x)$ on $L(\theta, b)$, respectively. Set

$$\begin{aligned} a_1 &= ax, & x_1 &= T_1(a_1) = a_1, & b_1 &= T_2(x_1), \\ a_2 &= [a - ax], & x_2 &= T_1(a_2), & b_2 &= T_2(x_2). \end{aligned}$$

Then $a = a_1 \oplus a_2$, $x = x_1 \oplus x_2$, $b = b_1 \oplus b_2$. By Lemma 3.3, $a_1 = x_1 \sim b_1$. Since the hypotheses of Lemma 4.5 and the special conditions of (β) are clearly satisfied by a_2, x_2, b_2, $a_2 \sim b_2$. Lemma 4.2 now gives $a \sim b$, and Lemma 4.5 is established for the special case (γ).

(δ) Suppose finally only the hypotheses of Lemma 4.5. The method by which (γ) was deduced from (β) can be applied in the same way to deduce (δ) from (γ). Thus Lemma 4.5 is proved.

COROLLARY. *If T_1, T_2 are perspective mappings of $L(\theta, a)$ on $L(\theta, x)$, and of $L(\theta, x)$ on $L(\theta, b)$, respectively, and if T_2T_1 maps a_1 on b_1, then $a_1b_1 = \theta$ implies $a_1 \sim b_1$.*

Proof. Set $x_1 = T_1(a_1)$. Lemmas 3.3 and 4.5 applied to a_1, x_1, b_1, give the desired result.

5. **Transitivity of perspectivity.** We prove the following lemma:

LEMMA 5.1. *If $a_1 \geqq a_2 \geqq \cdots$ and c are given, and if $\theta = \prod_n (a_n c)$, then there exist decompositions*

$$a_n = a_n c \oplus a_n' ,$$

for $n = 1, 2, \cdots$, such that $a_1' \geqq a_2' \geqq \cdots$.

Proof. Let $I_n = [a_n - (a_n c + a_{n+1})]$, for $n = 1, 2, \cdots$, and let $\bar{a} = \prod_n a_n$, $\bar{I} = [\bar{a} - \bar{a}c]$. Then

$$\begin{aligned} a_1 &= I_1 \oplus (a_1 c + a_2), \\ a_2 &= I_2 \oplus (a_2 c + a_3), \\ &\cdots\cdots\cdots, \\ a_n &= I_n \oplus (a_n c + a_{n+1}), \\ &\cdots\cdots\cdots, \\ \bar{a} &= \bar{I} \oplus \bar{a}c; \end{aligned}$$

and clearly $c\bar{I} = c\bar{I}\bar{a}c = \theta$, $cI_r = cI_r(a_r c + a_{r+1}) = \theta$, $I_{r+k} \leqq a_{r+k} \leqq a_r$, for all $r \geqq 1$, $k \geqq 0$.

We can now prove that $a_r = a_r c \oplus \bar{I} \oplus \sum_{n=r}^{\infty} I_n$, for $r = 1, 2, \cdots$. In the first place, $a_r c$, $\bar{I}$, I_n, $(n = r, r+1, \cdots)$ are independent over θ by the corollary to Lemma 2.6 since $a_r c \bar{I} = \theta$; and for all $p > n \geqq r$,

$$\begin{aligned} I_n\left(a_r c + \bar{I} + \sum_{m=n+1}^{p} I_m\right) &= I_n a_n (a_r c + a_{n+1})\left(a_r c + \bar{I} + \sum_{m=n+1}^{p} I_m\right) \\ &= I_n (a_n c + a_{n+1})\left(a_r c + \bar{I} + \sum_{m=n+1}^{p} I_m\right) \\ &= \theta. \end{aligned}$$

Secondly,

$$\begin{aligned} a_r c + \bar{I} + \sum_{n=r}^{\infty} I_n &= a_r c + \prod_{m=r}^{\infty} a_m + \sum_{n=r}^{\infty} I_n \\ &= \prod_{m=r}^{\infty}\left(a_r c + a_m + \sum_{n=r}^{\infty} I_n\right) \\ &= \prod_{m=r}^{\infty}\left(a_r c + a_m + \sum_{n=r}^{m-1} I_n\right). \end{aligned}$$

Now if $m \geqq r$,

$$a_r c + a_m + \sum_{n=r}^{m-1} I_n = a_r c + a_{m-1} c + a_m + I_{m-1} + \sum_{n=r}^{m-2} I_n$$
$$= a_r c + a_{m-1} + \sum_{n=r}^{m-2} I_n = \cdots$$
$$= a_r c + a_r = a_r.$$

Thus $a_r c \oplus I \oplus \sum_{n=r}^{\infty} I_n = \prod_{m=r}^{\infty}(a_r) = a_r$ as required. It is now clear that if we set

$$a_n' = I \oplus \sum_{r=n}^{\infty} I_r,$$

we will obtain the decompositions required by the lemma.

COROLLARY. *In Lemma* 5.1, I, I_n, $(n=1, 2, \cdots)$, *are independent over* θ.

Proof. This is immediate from the proof of Lemma 5.1.

THEOREM 5.1. TRANSITIVITY OF PERSPECTIVITY. *$a\sim x$, $x\sim b$ together imply $a\sim b$.*

Proof. (I) Let $\theta = abx$, and let T_1, T_2 be perspective mappings of $L(\theta, a)$ on $L(\theta, x)$ and of $L(\theta, x)$ on $L(\theta, b)$, respectively. Let $T = T_2 T_1$ be the product mapping of $L(\theta, a)$ on $L(\theta, b)$, and let T^{-1} be the inverse mapping to T. We shall use the notation $a_1 \to b_1$, or the equivalent $b_1 = T(a_1)$, to denote that b_1 is the map of a_1 under T.

(II) Let $c = ab$, and let $a_0 = [a - c]$, where a_0 is restricted to satisfy a certain condition which will be stated precisely later (see (III) below). Set $b_1 = (T(a_0))c$, $b_1' = [T(a_0) - b_1]$, $a_1 = T^{-1}(b_1)$, $a_1' = T^{-1}(b_1')$. Then

$$a = a_0 \oplus c,$$
$$a_0 = a_1 \oplus a_1', \qquad a_1 \to b_1 \leqq c, \qquad a_1' \to b_1', \qquad (b_1' c = \theta).$$

Since $\theta \leqq b_1 \leqq a$, $T(b_1)$ is defined. Let $\{T(b_1)\}c = b_2$, $b_2' = [T(b_1) - b_2]$, $b_{12} = T^{-1}(b_2)$, $a_2 = T^{-1}(b_{12})$, $b_{12}' = T^{-1}(b_2')$, $a_2' = T^{-1}(b_{12}')$. Then

$$a_1 = a_2 \oplus a_2', \qquad a_2 \to b_{12} \to b_2 \leqq c, \qquad a_2' \to b_{12}' \to b_2', \qquad (b_2' c = \theta).$$

Similarly, obtain the table

$$\begin{aligned}
a_0 &= a_1 \oplus a_1', && a_1 \to b_1 \leqq c, && a_1' \to b_1', && (b_1' c = \theta),\\
a_1 &= a_2 \oplus a_2', && a_2 \to b_{12} \to b_2 \leqq c, && a_2' \to b_{12}' \to b_2', && (b_2' c = \theta),\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots, &&&&&&\\
a_{n-1} &= a_n \oplus a_n', && a_n \to b_{12\ldots n} \to b_{23\ldots n} \to \cdots \to b_n \leqq c, &&&&\\
& && a_n' \to b_{12\ldots n}' \to b_{23\ldots n}' \to \cdots \to b_n', && && (b_n' c = \theta),\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots. &&&&&&
\end{aligned}$$

We can now prove the following statements:

(α) $\prod_{n-1}^{\infty} a_n = \theta$ *and* $a_0 = \sum_{n-1}^{\infty}(\oplus a_n')$.

(β) a_n', $b_{12\cdots n}'$, $b_{23\cdots n}'$, $\cdots$, b_n' *are independent over* θ, *for* $n = 1, 2, \cdots$.

(γ) *If we set* $d_n = a_n' + b_{12\cdots n}' + b_{23\cdots n}' + \cdots + b_n'$, *then* $d_1, d_2, \cdots$ *are independent over* θ.

(δ) *All the primed elements* a_1', a_2', $\cdots$, b_1', b_{12}', $\cdots$ *are independent over* θ.

Proof of (α). Let $\prod_n a_n = \bar{a}$. Then $T^n(\bar{a}) = TT \cdots T(\bar{a})$ (n factors T) is defined for $n = 1, 2, \cdots$; and $T(\bar{a})$, $T^2(\bar{a})$, $\cdots$ are independent over θ by the corollary to Lemma 2.6 since

$$\{T^n(\bar{a})\}\{T^{n+1}(\bar{a}) + \cdots + T^{n+p}(\bar{a})\}$$

has a T^{-n} map which is $\leqq a_0 b = a_0 a b = a_0 c = \theta$, for all n, $p \geqq 1$. Since $T^n(\bar{a}) \sim T^{n+1}(\bar{a})$ by Lemma 3.3, Lemma 4.3 shows that $\bar{a} = \theta$. The statement (α) now follows from Lemma 2.11.

Proof of (β).

$$b_{r(r+1)\cdots n}'(b_{(r+1)(r+2)\cdots n}' + b_{(r+2)(r+3)\cdots n}' + \cdots + b_n') = \theta$$

since it has a T^{-r} map which is $\leqq a_0 b = \theta$. The statement (β) now follows from the corollary to Lemma 2.6.

Proof of (γ). $\theta \leqq d_n(d_{n+1} + d_{n+2} + \cdots + d_{n+p})$

$$\leqq (a_n' + b_{12\cdots n}' + \cdots + b_n')\left\{\begin{array}{r} a_{n+1}' + b_{12\cdots(n+1)}' + \cdots + b_{n+1}' \\ + a_{n+2}' + b_{12\cdots(n+2)}' + \cdots + b_{n+2}' \\ + \cdots\cdots\cdots\cdots \\ + a_{n+p}' + b_{12\cdots(n+p)}' + \cdots + b_{n+p}' \end{array}\right\}$$

$$= a_n'(a_{n+1}' + a_{n+2}' + \cdots + a_{n+p}') + (b_{12\cdots n}' + \cdots + b_n')\left\{\begin{array}{r} b_{12\cdots(n+1)}' + \cdots + b_{n+1}' \\ + b_{12\cdots(n+2)}' + \cdots + b_{n+2}' \\ + \cdots\cdots\cdots\cdots \\ + b_{12\cdots(n+p)}' + \cdots + b_{n+p}' \end{array}\right\}$$

$$= \theta + T\left\{(a_n + b_{12\cdots n} + \cdots + b_{(n-1)n})\left\{\begin{array}{r} a_{n+1}' + b_{12\cdots(n+1)}' + \cdots + b_{n(n+1)}' \\ + a_{n+2}' + b_{12\cdots(n+2)}' + \cdots + b_{(n+1)(n+2)}' \\ + \cdots\cdots\cdots\cdots\cdots\cdots \\ + a_{n+p}' + b_{12\cdots(n+p)}' + \cdots + b_{(n+p-1)(n+p)}' \end{array}\right\}\right\}$$

$$\leqq \cdots \leqq T^{n+1}(\theta) = \theta.$$

The statement (γ) now follows from the corollary to Lemma 2.6.

Proof of (δ). The statement (δ) follows from (β) and (γ) by Lemma 2.10.

(III) By the corollary to Lemma 4.5 $a_n' \sim b_{12\cdots n}'$, $b_{12\cdots n}' \sim b_{23\cdots n}'$, $\cdots$, $b_{(n-1)n} \sim b_n'$. Since a_n', $b_{12\cdots n}'$, $\cdots$, b_n' are independent, repeated application

of Lemma 4.1 shows that $a_n' \sim b_n'$ for $n = 1, 2, \cdots$. Now set

$$A = \sum_{n=1}^{\infty} (\oplus\, a_n') \oplus \sum_{r \neq s} (\oplus\, b'_{r(r+1)\cdots s}),$$

$$B = \sum_{n=1}^{\infty} (\oplus\, b_n') \oplus \sum_{r \neq s} (\oplus\, b'_{r(r+1)\cdots s}).$$

By Lemma 4.2, $A \sim B$. Furthermore $T(A) = B$.

Now suppose that a_0 was chosen (in (II) above) in such a way that, the b_n' having been defined as above, we should have $(\sum_{n=1}^{\infty} b_n')c = \theta$ (that such an a_0 exists will be shown in (V) below). Setting

$$g = \left[c - \sum_{r \neq s} (\oplus\, b'_{r(r+1)\cdots s}) \right]$$

we have $Bg = Bgc = g \sum_{r \neq s} (\oplus b'_{r(r+1)\cdots s}) = \theta$, and we can define

$$h = [b - (B \oplus g)].$$

Then we clearly have

$$c = \sum_{r \neq s} (\oplus\, b'_{r(r+1)\cdots s}) \oplus g,$$

$$a = A \oplus g,$$

$$b = B \oplus g \oplus h.$$

By Lemma 4.2, $a \sim b$ if only $h = \theta$.

(IV) We proceed to show that $h = \theta$. Let $g' = T^{-1}(g + h)$. Since

$$A \oplus g = a = T^{-1}(b) = T^{-1}(B \oplus g \oplus h) = A \oplus g',$$

it follows that there exists a perspective mapping S of $L(\theta, g')$ on $L(\theta, g)$. Now set $h_0 = h$ and define h_n', h_n, for $n = 1, 2, \cdots$, by induction on n as follows:

$$h_n' = T^{-1}(h_{n-1}), \qquad h_n = S(h_n').$$

Then $h_0, h_1, \cdots$ are independent over θ by the corollary to Lemma 2.6 since

$$\begin{aligned} h_n(h_{n+1} + \cdots + h_{n+p}) &= (ST^{-1})^n \{ h(h_1 + \cdots + h_p) \} \\ &= (ST^{-1})^n \{ hg(h_1 + \cdots + h_p) \} \\ &= (ST^{-1})^n(\theta) = \theta. \end{aligned}$$

Since $h_{n-1} \sim h_n'$, $h_n' \sim h_n$, and $h_{n-1}h_n = \theta \leqq h_n'$, Lemma 4.5 shows that $h_{n-1} \sim h_n$. Now Lemma 4.3 gives $h = h_0 = \theta$ as required.

(V) To complete the proof of Theorem 5.1 we need only show that the $a_0 = [a - ab] = [a - c]$ defined in (II) above could be chosen in such a way that $c\sum_{n=1}^{\infty} b_n' = \theta$ will hold in (III) above. We first note that if we set

$$v_1 = T^{-1}(c),\ v_2 = T^{-1}(v_1 c),\ \cdots,\ v_{n+1} = T^{-1}(v_n c),\ \cdots,$$

then it is sufficient to choose a_0 so that $a = a_0 \oplus c$ and

$$v_n c = (v_n + a_0)c,$$

for $n = 1, 2, \cdots$. For if a_0 is so chosen, then

$$\begin{aligned}
\theta &\leqq b_{n+1}'(c + b_1' + b_2' + \cdots + b_n') \\
&= T\{b_{n(n+1)}'(T^{-1}(c) + a_1' + b_{12}' + \cdots + b_{(n-1)n}')\} \\
&\leqq T\{b_{n(n+1)}' c(v_1 + a_0 + b_{12}' + \cdots + b_{(n-1)n}')\} \\
&= T\{b_{n(n+1)}'(c(v_1 + a_0) + b_{12}' + \cdots + b_{(n-1)n}')\} \\
&= T\{b_{n(n+1)}'(v_1 c + b_{12}' + \cdots + b_{(n-1)n}')\} \\
&\leqq T^2\{b_{(n-1)n(n+1)}'(v_2 + a_0 + b_{123}' + \cdots + b_{(n-2)(n-1)n}')\} \\
&\leqq T^2\{b_{(n-1)n(n+1)}'(v_2 c + b_{123}' + \cdots + b_{(n-2)(n-1)n}')\} \\
&\leqq \cdots \leqq T^n(b_{12\cdots(n+1)}' v_n c) \leqq b_{n+1}' c = \theta;
\end{aligned}$$

hence

$$b_{n+1}'(c + b_1' + b_2' + \cdots + b_n') = \theta,$$

for $n = 1, 2, \cdots$. By Lemma 2.6, c, b_1', b_2', $\cdots$ are independent over θ; hence $c\sum_{n=1}^{\infty} b_n' = \theta$ as required.

Thus we have only to construct an a_0 such that

$$a = a_0 \oplus c, \quad v_n c = (v_n + a_0)c,$$

for $n = 1, 2, \cdots$. Apply Lemma 5.1 to $v_1 \geqq v_2 \geqq \cdots$ and c, and obtain I, I_n, $n = 1, 2, \cdots$, as in the proof of Lemma 5.1. Then $v_n = v_n c \oplus I \oplus \sum_{m=n}^{\infty} I_m$. Let $u = [a - (v_1 + c)]$. Then

$$u\left(I + \sum_{m=1}^{\infty} I_m\right) = u v_1\left(I + \sum_{m=1}^{\infty} I_m\right) = \theta,$$

and we can set

$$a_0 = I \oplus \sum_{m=1}^{\infty} (\oplus I_m) \oplus u.$$

This a_0 satisfies our requirements, for

$$a_0 + c = c + I + \sum_{m=1}^{\infty} I_m + u = c + v_1 + u = a,$$

$$a_0c = a_0(v_1 + c)c = \left(I + \sum_{m=1}^{\infty} I_m\right)c$$
$$= \left(I + \sum_{m=1}^{\infty} I_m\right)v_1c = \theta.$$

Hence $a_0 \oplus c = a$. And

$$(v_n + a_0)c = \left(v_nc + I + \sum_{m=1}^{\infty} I_m + u\right)c$$
$$= (v_nc + a_0)c$$
$$= v_nc + a_0c = v_nc + \theta = v_nc.$$

This completes the proof of Theorem 5.1.

6. **Additivity and continuity properties of perspectivity.** We prove the following lemma:

LEMMA 6.1. *If θ is defined and if*

$$a = a_1 \oplus a_1' = a_2 \oplus a_2',$$

then $a_1 \sim a_2$ implies $a_1' \sim a_2'$.

Proof. The perspectivity $a_1 \sim a_2$ implies, by Lemma 3.1, the existence of an x for which

$$a_1 + x = a_2 + x = a_1 + a_2,$$
$$a_1x = a_2x = \theta.$$

Let $c = [a - (a_1 + a_2)]$. Then a_1' is perspective to $(x+c)$ with axis a_1, for we have the relation

$$a_1' + a_1 = a = c + (a_1 + a_2) = c + (a_1 + x)$$
$$= (x + c) + a_1,$$

hence relation (i) holds; and

$$a_1'a_1 = \theta = xa_1 = (x + \theta)a_1 = \{x + (a_1 + a_2)c\}a_1$$
$$= \{x + (a_1 + x)c\}a_1 = (x + c)a_1,$$

hence relation (ii) holds.

Similarly a_2' is perspective to $(x+c)$. Theorem 5.1 then proves that $a_1' \sim a_2'$ as required.

LEMMA 6.2. *If θ is defined, and if*

$$a = a_1 \oplus a_1', \qquad b = b_1 \oplus b_1',$$

then $a \sim b$, $a_1 \sim b_1$ together imply $a_1' \sim b_1'$.

Proof. Let T be a perspective mapping of $L(\theta, a)$ on $L(\theta, b)$, and let $u=T(a_1)$, $v=T(a_1')$; then $a_1\sim u$, $a_1'\sim v$, and $b=u\oplus v$. Since $a_1\sim b_1$, Theorem 5.1 gives $u\sim b_1$ and Lemma 6.1 gives $v\sim b_1'$. Since $a_1'\sim v$ and $v\sim b_1'$, Theorem 5.1 gives $a_1'\sim b_1'$.

LEMMA 6.3. *$a_1\sim b_1$, $a_2\sim b_2$, $a_1a_2=b_1b_2$ together imply $(a_1+a_2)\sim(b_1+b_2)$.*

Proof. Let $\theta=a_1a_2=b_1b_2$, $d=a_1+a_2+b_1+b_2$, and define $a'=[d-(a_1\oplus a_2)]$, $b'=[d-(b_1\oplus b_2)]$. Then

$$(a'\oplus a_2)\oplus a_1 = d = (b'\oplus b_2)\oplus b_1.$$

By Lemma 6.1, $(a'\oplus a_2)\sim(b'\oplus b_2)$; hence by Lemma 6.2 $a'\sim b'$. Since $a'\oplus(a_1\oplus a_2)=b'\oplus(b_1\oplus b_2)$ Lemma 6.1 proves $(a_1+a_2)\sim(b_1+b_2)$ as required.

LEMMA 6.4. *If θ is defined, and if $a_1, \cdots, a_p$ and $b_1, \cdots, b_p$ are two sets of elements, each independent over θ, with $a_r\sim b_r$ for $r=1, \cdots, p$, where $p=1, 2, \cdots$, then*

$$\left(\sum_{r=1}^{p}(\oplus\, a_r)\right)\sim\left(\sum_{r=1}^{p}(\oplus\, b_r)\right).$$

Proof. Suppose the lemma established for $p=n$ for some fixed $n=1, 2, \cdots$. Then

$$(a_1\oplus\cdots\oplus a_n)\sim(b_1\oplus\cdots\oplus b_n), \qquad a_{n+1}\sim b_{n+1}$$

imply, by Lemma 6.3, $(a_1+\cdots+a_n+a_{n+1})\sim(b_1+\cdots+b_n+b_{n+1})$; and the lemma will hold for $p=n+1$. Since the lemma is trivially true for $p=1$, it holds, by induction, for all $p=1, 2, \cdots$.

We now define a relation $a\propto b$ as follows:

DEFINITION 6.1. *$a\propto b$ if $a\sim b_1$ for some $b_1\leqq b$.*

LEMMA 6.5. (I) *$a\leqq b$ implies $a\propto b$.*
(II) *$a\leqq b$, $b\propto c$ together imply $a\propto c$.*
(III) *$a\propto b$, $b\propto c$ together imply $a\propto c$.*
(IV) *$a\propto b$, $b\propto a$ together imply $a\sim b$.*

Proof. (I) follows from $a\sim a$.

(II): Let $\theta=ac$, and let T be a perspective mapping of $L(\theta, b)$ on $L(\theta, c)$. Then $T(a)$ is defined, $a\sim T(a)$, and $T(a)\leqq c$. Hence $a\propto c$.

(III): $a\propto b$ means $a\sim b_1$ for some $b_1\leqq b$. Since $b\propto c$, (II) gives $b_1\sim c_1$ for some $c_1\leqq c$. Theorem 5.1 then gives $a\sim c_1$. Hence $a\propto c$ as required.

(IV): $a\propto b$ means $a\sim b_1$ for some $b_1\leqq b$. If $b\propto a$, then $b\sim a_1$ for some $a_1\leqq a$, and $b_1\sim a_2$ for some $a_2\leqq a_1$, by (II). Then, by Theorem 5.1, $a\sim a_2$; and,

since $a_2 \leqq a$, Lemma 4.4 implies $a_2 = a$. Since $a_2 \leqq a_1 \leqq a$, we have $a = a_1 \sim b$; that is $a \sim b$ as required.

LEMMA 6.6. *If $c \propto a$ and $d = ca$, there exists an element a_1 with $d \leqq a_1 \leqq a$ and $c \sim a_1$.*

Proof. $c \propto a$ means $c \sim a_2$ for some $a_2 \leqq a$. Let $\theta = a_2 d$; then $\theta \leqq ca_2$, and there exists, by Lemmas 3.1 and 3.2, a perspective mapping T of $L(\theta, c)$ on $L(\theta, a_2)$. Define $c' = [c - d]$ and $\bar{c}' = T(c')$; then $\bar{c}' \leqq a_2 \leqq a$, $c'd = \theta$, $c' \sim \bar{c}'$, and $\bar{c}'d = \bar{c}'a_2 d = \theta$. By Lemma 6.3,

$$c = (c' \oplus d) \sim (\bar{c}' \oplus d) \leqq a,$$

and $a_1 = \bar{c}' \oplus d$ satisfies all the requirements of the lemma.

LEMMA 6.7. *If θ is defined and if $a \oplus a_1 \geqq a' \oplus a_2$, then $a \sim a'$ implies $a_2 \propto a_1$.*

Proof. Let $v = [(a \oplus a_1) - (a' \oplus a_2)]$. Then $a \oplus a_1 = a' \oplus a_2 \oplus v$, and Lemma 6.1 implies $a_1 \sim (a_2 \oplus v)$; thus $a_2 \propto a_1$ as required.

LEMMA 6.8. *If θ is defined and if $a \oplus a_2 \propto a \oplus a_1$, then $a_2 \propto a_1$.*

Proof. By Lemma 6.6 (with the θ of the present lemma in the place of the d of Lemma 6.6) $(a \oplus a_2) \sim u$, where $\theta \leqq u \geqq (a + a_1)$. Let T be a perspective mapping of $L(\theta, a + a_2)$ on $L(\theta, u)$, and let $\bar{a} = T(a)$ and $\bar{a}_2 = T(a_2)$; then $a \oplus a_1 \geqq \bar{a} \oplus \bar{a}_2$ and $a \sim \bar{a}$. By Lemma 6.7, $a_2 \propto a_1$, and since $\bar{a}_2 \sim a_2$, Lemma 6.5 (III) implies $a_2 \propto \bar{a}_1$, which proves the lemma.

LEMMA 6.9. *If θ is defined, and if*

$$u \oplus v_1 \oplus \cdots \oplus v_p \propto a,$$

where $u \leqq a$ and $(v_1 \oplus \cdots \oplus v_p)a = \theta$, $(p = 1, 2, \cdots)$, then there exist elements $v_1', \cdots, v_p'$ with $v_r \sim v_r'$, $(r = 1, \cdots, p)$, $u, v_1, \cdots, v_p, v_1', \cdots, v_p'$ independent over θ, and $u \oplus v_1' \oplus \cdots \oplus v_p' \leqq a$.

Proof. Define $a_1 = [a - u]$; then

$$a = u \oplus a_1,\ u \oplus (v_1 \oplus \cdots \oplus v_p) \propto u \oplus a_1,$$

and Lemma 6.8 implies $(v_1 \oplus \cdots \oplus v_p) \propto a_1$. Thus there exists, by Lemma 6.6, a perspective mapping T of $L(\theta, (v_1 \oplus \cdots \oplus v_p))$ on $L(\theta, a_2)$ for some a_2 satisfying $\theta \leqq a_2 \leqq a_1$. Let $v_r' = T(v_r)$ for $r = 1, \cdots, p$. Then $v_r \sim v_r'$, for $r = 1, \cdots, p$; $\sum_{r=1}^{p} (\oplus v_r') \leqq a_2 \leqq a_1 \leqq a$; and v_r', $(r = 1, \cdots, p)$, are independent over θ. Since

$$\left(\sum_{r=1}^{p} (\oplus\, v_r')\right)\left(u \oplus \sum_{r=1}^{p} (\oplus\, v_r)\right) = \left(\sum_{r=1}^{p} (\oplus\, v_r')\right) a_1 a \left(u \oplus \sum_{r=1}^{p} (\oplus\, v_r)\right)$$

$$= \left(\sum_{r=1}^{p} (\oplus\, v_r')\right) a_1 u = \theta,$$

$u, v_1, \cdots, v_p, v_1', \cdots, v_p'$ are independent over θ by Lemma 2.10. Thus $v_1', \cdots, v_p'$ satisfy all the requirements of the lemma.

LEMMA 6.10. *Let $a_1 \geqq a_2 \geqq \cdots$ be an infinite set of elements, and let c satisfy $\theta \leqq c \leqq a_1$, where $\theta = \prod_n a_n$. Suppose further that $c \propto a_n$ for $n = 1, 2, \cdots$. Then there exists an element $c' \leqq a_1$ such that $c \sim c'$, $cc' = \theta$, and $(c \oplus c') \propto a_n$ for $n = 1, 2, \cdots$.*

Proof. Define $c_t = [ca_t - ca_{t+1}]$ for $t = 1, 2, \cdots$; then $ca_t = ca_{t+1} \oplus c_t$, and

$$c = ca_1 = \sum_{t=1}^{\infty} (\oplus\, c_t) \oplus \prod_{t=1}^{\infty} (ca_t)$$

$$= \sum_{t=1}^{\infty} (\oplus\, c_t) + \theta = \sum_{t=1}^{\infty} (\oplus\, c_t).$$

Suppose that $c_{r_1 r_2 \cdots r_n t}$ have been defined, for all $1 \leqq r_1 < r_2 < \cdots < r_n < t < \infty$ with $1 \leqq r_n < p$ for some fixed $p = 1, 2, \cdots$ (n taking all values possible), in such a way that the following conditions are satisfied:

$(\alpha)_p$ c_t, $t = 1, 2, \cdots$; $c_{r_1 r_2 \cdots r_n t}$, $1 \leqq r_1 < r_2 < \cdots < r_n < t$, for $r_n < p$, are independent over θ,

$(\beta)_p$ If we set

$$c'_{r_1 r_2 \cdots r_n} = \sum_{t=r_n+1}^{\infty} (\oplus\, c_{r_1 r_2 \cdots r_n t}),$$

then

$$c_{r_1 r_2 \cdots r_n t} = [c'_{r_1 r_2 \cdots r_n} a_t - c'_{r_1 r_2 \cdots r_n} a_{t+1}],$$

and

$$c_{r_1 r_2 \cdots r_n} \sim c'_{r_1 r_2 \cdots r_n}, \quad 1 \leqq r_1 < r_2 < \cdots < r_n < t < \infty, \ r_n < p.$$

$(\gamma)_p$ $$c \sim \sum_{t=p}^{\infty} (\oplus\, c_t) \oplus \sum_{t=p}^{\infty} (\oplus\, c_{r_1 r_2 \cdots r_n t}),$$

where in the last summation $r_1, r_2, \cdots, r_n$ take on all possible values with $r_n < p$.

Then $c \propto a_{p+1}$, and

$$c \sim \left(\sum_{t=p+1}^{\infty} (\oplus\, c_t) \oplus \sum_{t=p+1, r_n < p}^{\infty} (\oplus\, c_{r_1 r_2 \cdots r_n t}) \right)$$

$$\oplus \left(c_p \oplus \sum_{r_n < p} (\oplus\, c_{r_1 r_2 \cdots r_n p}) \right);$$

and Lemma 6.9 secures the existence of elements c_p', $c'_{r_1r_2\cdots r_np}$, $(1\leqq r_1<r_2<\cdots<r_n<p)$, such that $\theta\leqq c_p'\leqq a_{p+1}$, $\theta\leqq c_{r_1r_2\cdots r_np}\leqq a_{p+1}$, $c_p\sim c_p'$, $c_{r_1r_2\cdots r_np}\sim c'_{r_1r_2\cdots r_np}$, and c_t, $(t=1, 2, \cdots)$, $c_{r_1r_2\cdots r_nt}$, $(1\leqq r_1<r_2<\cdots<r_n<t<\infty$, $r_n<p)$, c_p', $c'_{r_1r_2\cdots r_np}$, $(1\leqq r_1<r_2<\cdots<r_n<p)$, are independent over θ. Hence if we define $c_{pt}=[c_p'a_t-c_p'a_{t+1}]$ and $c_{r_1r_2\cdots r_npt}=[c'_{r_1r_2\cdots r_np}a_t-c'_{r_1r_2\cdots r_np}a_{t+1}]$, then $(\alpha)_{p+1}$, $(\beta)_{p+1}$, and $(\gamma)_{p+1}$ will be satisfied. Since $(\alpha)_1$, $(\beta)_1$, and $(\gamma)_1$ are trivially satisfied, it follows that we can define the $c_{r_1r_2\cdots r_nt}$, (for all $1\leqq r_1<r_2<\cdots<r_n<t<\infty$), so that $(\alpha)_p$, $(\beta)_p$, and $(\gamma)_p$ are satisfied for all $p=1, 2, \cdots$. Then $c_1, c_2, \cdots, c_1', c_2', \cdots$ are independent over θ, and $c_n\sim c_n'$ for $n=1, 2, \cdots$. Lemmas 2.3 and 4.2 now imply that $c=\sum_{n=1}^{\infty}(\oplus c_n)\sim\sum_{n=1}^{\infty}(\oplus c_n')$. If we set $c'=\sum_{n=1}^{\infty}(\oplus c_n')$, we will have $c\sim c'$, $cc'=\theta$, and $(c\oplus c')\leqq a_1$.

Finally, since $c\propto a_p$, we have $c\sim c^p$ for some c^p with $\theta\leqq c^p\leqq a_p$ by Lemma 6.6. Applying the reasoning of the preceding paragraph to c^p and $a_p\geqq a_{p+1}\geqq\cdots$ we obtain $c^{p'}$ such that $c^p\sim c^{p'}$ and $(c^p\oplus c^{p'})\leqq a_p$. Then $c'\sim c$, $c\sim c^p$, $c^p\sim c^{p'}$, imply, by Theorem 5.1, $c'\sim c^{p'}$ and hence, by Lemma 6.3, $(c\oplus c')\sim(c^p\oplus c^{p'})$. Lemma 6.5 (III) now implies that $(c\oplus c')\propto a_p$ for all $p=1, 2, \cdots$, and c' satisfies all the requirements of the lemma.

LEMMA 6.11. *The hypotheses of Lemma* 6.10 *imply* $c=\theta$.

Proof. Suppose that $\bar{c}_1, \bar{c}_2, \cdots, \bar{c}_{2^p}$ have been defined for some fixed $p=0, 1, \cdots$ in such a way that, if we write $c_{(p)}$ for $\bar{c}_1+\cdots+\bar{c}_p$, then

$(\lambda)_p$ $c\sim\bar{c}_r$ for $r=1, \cdots, 2^p$;

$(\mu)_p$ $\bar{c}_r$, $r=1, \cdots, 2^p$, are independent over θ;

$(\nu)_p$ $c_{(p)}\leqq a_1$ and $c_{(p)}\propto a_n$ for $n=1, 2, \cdots$.

Then there exists, by Lemma 6.10, an element $c'\leqq a_1$ such that $c_{(p)}\sim c'$, $c_{(p)}c'=\theta$, and $(c_{(p)}\oplus c')\propto a_n$ for $n=1, 2, \cdots$. Let T be a perspective mapping of $L(\theta, c_{(p)})$ on $L(\theta, c')$, and let $\bar{c}_{2^p+r}=T(\bar{c}_r)$ for $r=1, \cdots, 2^p$. Then $(\lambda)_{p+1}$, $(\mu)_{p+1}$, and $(\nu)_{p+1}$ will be satisfied. Since we can define $\bar{c}_1=c$ to satisfy $(\lambda)_0$, $(\mu)_0$, and $(\nu)_0$, it follows that we can define, by induction on p, an infinite sequence $\bar{c}_n$, $(n=1, 2, \cdots)$, satisfying $(\lambda)_p$, $(\mu)_p$, and $(\nu)_p$, for all $p=0, 1, \cdots$. Then we have $\bar{c}_1, \bar{c}_2, \cdots$ independent over θ and $\bar{c}_n\sim\bar{c}_{n+1}$ (by Theorem 5.1 since $\bar{c}_n\sim c$, $c\sim\bar{c}_{n+1}$); hence $\bar{c}_1=\theta$ by Lemma 4.3. Since $c=\bar{c}_1$ we have $c=\theta$. This proves the lemma.

LEMMA 6.12. *Without the condition* $c\leqq a_1$, *the remaining hypotheses of Lemma* 6.10 *imply* $c=\theta$.

Proof. $\theta\leqq ca_1$, $c\propto a_1$, imply, by Lemma 6.6, the existence of a c_1 with $\theta\leqq c_1\leqq a_1$, and $c\sim c_1$. By Lemma 6.5 (III), c_1 and $a_1\geqq a_2\geqq\cdots$ satisfy the

hypotheses of Lemma 6.10; hence Lemma 6.11 implies $c_1=\theta$. Since $c\sim c_1=\theta$ and $\theta\leqq c$, Lemma 4.4 gives $c=\theta$. This proves the lemma.

LEMMA 6.13. *If θ is defined and $c\propto(a\oplus b)$, there exists a decomposition $c=c_1\oplus c_2$ with $c_1\propto a$, $c_2\propto b$.*

Proof. By Lemma 6.6, $c\sim u$, $\theta\leqq u\leqq(a\oplus b)$. Let $u_1=au$, define $u_1'=[u-u_1]$, and let $u_2=(a+u_1')b$; then $u=u_1\oplus u_1'$, and $u_2\sim u_1'$ with axis a, for

$$u_1'+a=(u_1'+a)(b+a)=(a+u_1')b+a=u_2+a,$$

hence (i) is satisfied; and

$$u_1'a=u_1'ua=u_1'u_1=\theta=u_2ba=u_2a,$$

hence (ii) is satisfied.

Now let $c_1=T^{-1}(u_1)$ and $c_2=T^{-1}(u_1')$; then $c=c_1\oplus c_2$, $c_1\sim u_1\leqq a$, and $c_2\sim u_2\leqq b$ (by Theorem 5.1, since $c_2\sim u_1'$ and $u_1'\sim u_2$). This proves the lemma.

LEMMA 6.14. *If $a_1\geqq a_2\geqq\cdots$ and $c\propto a_n$ for $n=1, 2, \cdots$, then $c\propto\prod_n a_n$.*

Proof. Let $\bar a=\prod_n a_n$, $\theta=c\bar a$. Apply Lemma 5.1 to $\bar a$ and $a_1\geqq a_2\geqq\cdots$ to obtain $a_n=a_n'\oplus\bar a$ with $a_1'\geqq a_2'\geqq\cdots$. Then $\prod_n a_n'=(\prod_n a_n')\bar a a_1'=\theta$.

Let $c_0=c$ and $\bar a_0=\bar a$. Suppose c_r, c_r', $\bar a_r$, $\bar a_r'$, have been defined for $1\leqq r<p$ and for some $p=1, 2, \cdots$ in such a way that the following conditions are satisfied:

$(\alpha)_p$ $c_{r-1}=c_r\oplus c_r'$, $\bar a_{r-1}=\bar a_r\oplus\bar a_r'$, $c_r'\sim\bar a_r'$, for $1\leqq r<p$.

$(\beta)_p$ $c_{p-1}\propto a_p'$, $(p>1)$, and $c_{p-1}\propto(a_n'+\bar a_{p-1})$, for $n=1, 2, \cdots$.

Then, since $c_{p-1}\propto(a_p'+\bar a_p)$, we can define c_p, c_p', $\bar a_p'$, by Lemma 6.13, so that

$$c_{p-1}=c_p\oplus c_p', \qquad c_p\propto a_p', \qquad c_p'\sim\bar a_p'\leqq\bar a_{p-1}.$$

Now define $\bar a_p=[\bar a_{p-1}-\bar a_p']$. Then, by the use of Lemma 6.8,

$$\bar a_{p-1}=\bar a_p\oplus\bar a_p', \qquad c_p\propto(a_n'\oplus\bar a_p), \quad \text{for} \quad n=1, 2, \cdots.$$

Thus $(\alpha)_{p+1}$, $(\beta)_{p+1}$ are satisfied. Since $(\alpha)_1$ and $(\beta)_1$ are satisfied by c_0, $\bar a_0$, it follows that we can define by induction c_r, c_r', $\bar a_r$, $\bar a_r'$, for $r=1, 2, \cdots$, to satisfy $(\alpha)_p$ and $(\beta)_p$ for all $p=1, 2, \cdots$.

By Lemma 2.11

$$c=\sum_{n=1}^{\infty}(\oplus c_n')\oplus\prod_{n=1}^{\infty}c_n, \qquad \bar a=\sum_{n=1}^{\infty}(\oplus\bar a_n')\oplus\prod_{n=1}^{\infty}\bar a_n.$$

Since

$$\prod_{n=1}^{\infty}a_n'=\theta\leqq\prod_{n=1}^{\infty}c_n\leqq c_r\propto a_r' \quad \text{for} \quad r=1, 2, \cdots,$$

Lemma 6.12 implies $\prod_{n=1}^{\infty} c_n = \theta$. Hence $c = \sum_{n=1}^{\infty}(\oplus c_n')$. Since

$$c\left(\sum_{n=1}^{\infty}(\oplus\, \bar{a}_n')\right) = c\bar{a}\left(\sum_{n=1}^{\infty}(\oplus\, \bar{a}_n')\right) = \theta,$$

Lemma 2.10 implies c_n', a_n', $(n=1, 2, \cdots)$, are independent. Lemmas 2.3 and 4.2 now give

$$c = \sum_{n=1}^{\infty}(\oplus\, c_n') \sim \sum_{n=1}^{\infty}(\oplus\, \bar{a}_n) \leqq \bar{a} = \prod_{n=1}^{\infty} a_n.$$

Hence $c \propto \prod_n a_n$, which proves the lemma.

LEMMA 6.15. *If θ is defined, and if*

$$a \oplus a' = b \oplus b',$$

then $b \propto a$ implies $a' \propto b'$.

Proof. Suppose $b \sim a_1 \leqq a$, and define $a_1' = [a - a_1]$. Then $a_1 \oplus (a_1' \oplus a') = b \oplus b'$, and Lemma 6.2 implies $(a_1' \oplus a') \sim b'$; hence by Lemma 6.5 (II) $a' \propto b'$.

LEMMA 6.16. *If $a_1 \leqq a_2 \leqq \cdots$ and if $a_n \propto c$ for $n = 1, 2, \cdots$, then $\sum_n a_n \propto c$.*

Proof. Let $a_1 c = \theta$, and define $u_1 = a_1$, $u_n = [a_n - a_{n-1}]$ for $n = 2, 3, \cdots$; then $a_n = \sum_{r=1}^{n} u_r$, $(n = 1, 2, \cdots)$, and $u_1, u_2, \cdots$ are independent over θ by Lemma 2.6, since $\bar{u}_{n+1}(u_1 + \cdots + u_n) = [a_{n+1} - a_n]a_n = \theta$ for $n = 1, 2, \cdots$. Set $b_n = \sum_{r=n+1}^{\infty}(\oplus u_r)$ for $n = 0, 1, \cdots$; then $b_0 \geqq b_1 \geqq \cdots$, and

$$b_0 = \sum_{r=1}^{\infty}(\oplus\, u_r) = a_n \oplus b_n, \qquad n = 1, 2, \cdots.$$

Hence $b_0 \geqq \sum_{n=1}^{\infty} a_n$. Since $u_n \leqq a_n$ we also have $b_0 = \sum_{n=1}^{\infty} u_n \leqq \sum_{n=1}^{\infty} a_n$; thus $b_0 = \sum_{n=1}^{\infty} a_n$. Now define $c' = [(b_0 + c) - c]$ and $b_0' = [(b_0 + c) - b_0]$. Then

$$c \oplus c' = b_0 \oplus b_0' = a_n \oplus (b_n \oplus b_0')$$

for $n = 1, 2, \cdots$, and Lemma 6.15 implies $c' \propto (b_n \oplus b_0')$ for $n = 1, 2, \cdots$. Applying Lemma 6.14 to c' and $(b_1 + b_0') \geqq (b_2 + b_0') \geqq \cdots$, we obtain

$$c' \propto \prod_n (b_n + b_0') = \left(\prod_n b_n\right) + b_0' = \theta + b_0' = b_0'.$$

Lemma 6.15 now implies $\sum_n a_n = b_0 \propto c$. This proves the lemma.

DEFINITION 6.2. *If $a_1, a_2, \cdots$ is an infinite sequence, we define*

$$\limsup a_n = \prod_p\left(\sum_{n=p}^{\infty} a_n\right), \qquad \liminf a_n = \sum_p\left(\prod_{n=p}^{\infty} a_n\right).$$

The sequence is called convergent if $\limsup a_n = \liminf a_n$, *and for a convergent sequence we define* $\lim a_n = \limsup a_n = \liminf a_n$.

LEMMA 6.17. *If* $a_1 \leqq a_2 \leqq \cdots$ $(a_1 \geqq a_2 \geqq \cdots)$ *then* $\lim a_n$ *is defined and is equal to* $\sum_n a_n$ $(\prod_n a_n)$.

Proof.

$$\limsup a_n = \prod_{p=1}^{\infty}\left(\sum_{n=p}^{\infty} a_n\right) = \prod_{p=1}^{\infty}\left(\sum_{n=1}^{\infty} a_n\right) = \sum_{n=1}^{\infty} a_n$$

$$\left(= \prod_{p=1}^{\infty}(a_p) = \prod_n a_n\right);$$

$$\liminf a_n = \sum_{p=1}^{\infty}\left(\prod_{n=p}^{\infty} a_n\right) = \sum_{p=1}^{\infty}(a_p) = \sum_n a_n$$

$$\left(= \sum_{p=1}^{\infty}\left(\prod_{n=1}^{\infty} a_n\right) = \prod_{n=1}^{\infty} a_n\right).$$

Hence $\limsup a_n = \liminf a_n = \sum_n a_n$ $(= \prod_n a_n)$ and the lemma follows from Definition 6.2.

THEOREM 6.1. CONTINUITY OF PERSPECTIVITY. *If* $a_1, a_2, \cdots$ *and* $b_1, b_2, \cdots$ *are convergent sequences with* $\lim a_n = \bar{a}$ *and* $\lim b_n = \bar{b}$, *then* $a_n \sim b_n$ *for* $n = 1, 2, \cdots$ *implies* $\bar{a} \sim \bar{b}$.

Proof. For every fixed $p = 1, 2, \cdots$ we have

$$\left(\prod_{n=p}^{\infty} a_n\right) \leqq a_r \sim b_r \leqq \left(\sum_{n=r}^{\infty} b_n\right), \qquad r = p, p+1, \cdots,$$

and Lemma 6.14 implies $(\prod_{n=p}^{\infty} a_n) \propto \prod_{r=p}^{\infty}(\sum_{n=r}^{\infty} b_n) = \bar{b}$. Lemma 6.16 gives $\bar{a} = \sum_{p=1}^{\infty}(\prod_{n=p}^{\infty} a_n) \propto \bar{b}$, that is $\bar{a} \propto \bar{b}$. Similarly $\bar{b} \propto \bar{a}$. Then, by Lemma 6.15(IV), $\bar{a} \sim \bar{b}$, which proves the theorem.

COROLLARY. *If* $a_1 \leqq a_2 \leqq \cdots$ *and* $b_1 \leqq b_2 \leqq \cdots$ $(a_1 \geqq a_2 \geqq \cdots$ *and* $b_1 \geqq b_2 \geqq \cdots)$ *then* $a_n \sim b_n$ *for* $n = 1, 2, \cdots$ *implies* $\sum_n a_n \sim \sum_n b_n$ $(\prod_n a_n \sim \prod_n b_n)$.

Proof. By Lemma 6.17 this is a special case of Theorem 6.1.

THEOREM 6.2. ADDITIVITY OF PERSPECTIVITY. *If* θ *is defined and if* a_n, $1 \leqq n < p$, *and* b_n, $1 \leqq n < p$, *are each independent over* θ, *where* p *is finite or infinite, then* $a_n \sim b_n$ *for* $1 \leqq n < p$ *implies*

$$\sum_{n=1}^{p}(\oplus a_n) \sim \sum_{n=1}^{p}(\oplus b_n).$$

Proof. By Lemma 6.4 $\sum_{n=1}^{r}(\oplus a_n) \sim \sum_{n=1}^{r}(\oplus b_n)$ for all $r < p$. If p is finite this proves the theorem, and if p is infinite we have, using the corollary to Theorem 6.1,

$$\sum_{n=1}^{\infty}(\oplus a_n) = \sum_{r}\left(\sum_{n=1}^{r}(\oplus a_n)\right) \sim \sum_{r}\left(\sum_{n=1}^{r}(\oplus b_n)\right) = \sum_{n=1}^{\infty}(\oplus b_n).$$

This proves the theorem.*

* For the special case of an irreducible geometry (finite dimensional or continuous), all the lemmas and theorems of §6 are easy consequences of the existence of a dimension function, and conversely, some of them are useful in establishing the existence of the dimension function (see C.G. part 1, chaps. 6 and 7). The notion of a convergent sequence is given in an equivalent form by von Neumann, Proceedings of the National Academy of Sciences, vol. 22 (1936), p. 107 (see the definition of lim** given there).

YALE UNIVERSITY,
NEW HAVEN, CONN.

PERMANENT CONFIGURATIONS IN THE PROBLEM OF FIVE BODIES*

BY

W. L. WILLIAMS

1. **Introduction.** In 1772 Lagrange found that if three bodies of arbitrary masses are placed at the vertices of an equilateral triangle and given the proper initial velocities, they will continue to be at the vertices of an equilateral triangle which may change in size, but not in shape. He also showed that if the three bodies are placed in a straight line and started off suitably, they will remain in a straight line, and the ratios of the mutual distances will be constant. More recently Moulton† has extended the straight line configuration to n bodies, and some particular instances of other configurations have been given by Hoppe,‡ Andoyer,§ and Longley.‖

A permanent configuration is a configuration of n bodies which has the property that the ratio of distances between corresponding bodies is constant. In other words, the figure may change in size, but not in shape. The object of this paper is to determine necessary and sufficient conditions for any plane configuration of five bodies, other than a straight line configuration, to be permanent and to give a detailed analysis of the various types of such configurations, both convex and concave, that can be formed by rearranging the bodies to give pentagons of different shapes. This work may be regarded as an extension of that of MacMillan and Bartky¶ in which the problem of four bodies is treated.

2. **The equations of motion.** Let $m_1, m_2, \cdots, m_5$ with coordinates (x_1,y_1), $(x_2, y_2), \cdots, (x_5, y_5)$, respectively, represent the masses of a plane system of five bodies referred to a set of rectangular axes, with origin at their common center of mass, which rotates with uniform angular velocity ω. If we assume that the bodies attract each other according to the Newtonian law and move in circles around the origin with uniform angular velocity ω, the differential

---

* Presented to the Society, September 6, 1938; received by the editors May 20, 1937.

† *Periodic Orbits*, Carnegie Institution, 1920, p. 285.

‡ *Erweiterung der bekannten Speciallösung des Dreikörperproblems*, Archiv der Mathematik und Physik, vol. 64, p. 218.

§ *Sur l'équilibre relatif de n corps*, Bulletin Astronomique, vol. 23 (1906), p. 50.

‖ *Some particular solutions in the problem of n bodies*, Bulletin of the American Mathematical Society, vol. 13 (1906–1907), p. 324.

¶ *Permanent configurations in the problem of four bodies*, these Transactions, vol. 34 (1932), pp. 838–875.

equations of motion reduce to

$$\omega^2 x_i = \sum_{j=1}^{5} \frac{m_j(x_i - x_j)}{r_{i,j}^3}, \qquad \omega^2 y_i = \sum_{j=1}^{5} \frac{m_j(y_i - y_j)}{r_{i,j}^3}, \qquad i = 1, 2, \cdots, 5; j \neq i, \tag{1}$$

where $r_{i,j}$ is the distance between the masses m_i and m_j.

It is to be observed that the equations

$$\sum_{j=1}^{5} m_i x_i = 0, \qquad \sum_{i=1}^{5} m_i y_i = 0, \tag{2}$$

which express the fact that the origin of coordinates is at the center of mass, follow directly from (1).

Define a constant r_0 by

$$\omega^2 = \sum_{i=1}^{5} m_i / r_0^3 . \tag{3}$$

The constant r_0 is to replace ω and will be used later in defining the regions for positive masses.

The equations (1), as a consequence of (2) and (3), may now be written in the equivalent form

$$\sum_{j=1}^{5} S_{i,j}(x_i - x_j)m_j = 0, \qquad \sum_{j=1}^{5} S_{i,j}(y_i - y_j)m_j = 0, \qquad i = 1, 2, \cdots, 5; j \neq i, \tag{4}$$

where

$$S_{i,j} = S_{j,i} = R_{i,j} - R_0, \quad R_{i,j} = R_{j,i} = 1/r_{i,j}^3, \quad R_0 = 1/r_0^3.$$

If in (4) m_i and m_{i+1}, $(i = 1, 2, \cdots, 5)$, are eliminated, respectively, from the $(i+1)$th and ith equations in x and the corresponding equations in y, and if we denote twice the area of the triangle whose vertices have as coordinates any three of the points (x_i, y_i), $(i = 1, 2, \cdots, 5)$, by D with a pair of numbers (from 1 to 5) as subscripts which do not occur as subscripts in the coordinates of the vertices of the triangle, we have

$$\begin{aligned} &S_{i,i+3}D_{i+1,i+2}m_i + S_{i+1,i+3}D_{i,i+2}m_{i+1} + S_{i+2,i+3}D_{i,i+1}m_{i+2} = 0, \\ &S_{i,i+4}D_{i+1,i+2}m_i + S_{i+1,i+4}D_{i,i+2}m_{i+1} + S_{i+2,i+4}D_{i,i+1}m_{i+2} = 0, \end{aligned} \tag{5}$$

or

$$
\begin{aligned}
& m_i/(S_{i+1,i+3}S_{i+2,i+4} - S_{i+1,i+4}S_{i+2,i+3})D_{i,i+2}D_{i,i+1} \\
(6) \qquad & = m_{i+1}/(S_{i+2,i+3}S_{i,i+4} - S_{i+2,i+4}S_{i,i+3})D_{i+1,i+2}D_{i,i+1} \\
& = m_{i+2}/(S_{i,i+3}S_{i+1,i+4} - S_{i,i+4}S_{i+1,i+3})D_{i+1,i+2}D_{i,i+2}, \\
& \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad i = 1, 2, \cdots, 5,
\end{aligned}
$$

a system of equations, equivalent to (4), in which the variables x and y do not occur explicitly. It is to be understood here as well as hereafter that the subscripts 6, 7, 8, and 9 are to be replaced by 1, 2, 3, and 4, respectively.

For each value of i in equations (6), two independent relations may be formed, each expressing a mass ratio as a function of the D's and S's. Among these ten mass ratios six distinct equalities exist; and when (6) is substituted in them we obtain, after some reduction,

$$
(7) \qquad \frac{S_{i,i+3}S_{i+1,i+2} - S_{i,i+2}S_{i+1,i+3}}{S_{i,i+2}S_{i+1,i+3} - S_{i,i+1}S_{i+2,i+3}} = \lambda_{i+4}, \qquad i = 1, 2, \cdots, 5,
$$

where

$$
(8) \qquad \lambda_{i+4} = \frac{D_{i+2,i+3}D_{i,i+1}}{D_{i+1,i+2}D_{i,i+3}}.
$$

It is to be observed that the subscript on λ corresponds to the missing subscript in the left-hand members of (7).

One may inquire whether the denominators in (6) may not vanish under certain conditions. That this is impossible may be seen by considering the several possible cases.

It should be stated first that in the five-body problem the masses are all different from zero. The form of equations (1) shows that this hypothesis has been made. Suppose, now, that no three of the bodies are in the same straight line (so that no D is zero), and let one of the expressions in S, say $S_{24}S_{35} - S_{25}S_{34}$, vanish. It follows at once from (6) that

$$
(A) \qquad S_{i,i+1}S_{i+2,i+4} = S_{i,i+2}S_{i+1,i+4} = S_{i,i+4}S_{i+1,i+2}, \qquad i = 1, 2, \cdots, 5.
$$

That is, if one denominator in (6) vanishes, all must vanish. But it is impossible to satisfy all of the equations (A) simultaneously, for if it were, then equations (5) would show that a permanent configuration of five bodies is obtained in which the five masses are arbitrary. Let two of these masses approach zero. The result, in so far as finite masses are concerned, is a permanant configuration of three bodies with arbitrary masses. The only such figure is an equilateral triangle. Hence, if the equations (A) are satisfied by a pentagon configuration, any three vertices of this pentagon must form an equilateral triangle. This is geometrically impossible.

Consider next the case where three or more of the bodies are in the same straight line. The case in which they are all in a line has been treated by Moulton, and the method of this paper is not applicable. Outside of Moulton's case there are only two five-body configurations in which some of the denominators in (6) vanish; namely, the square and the rhombus with one mass at each vertex and the fifth mass at the center. These two cases are treated in Examples 2 and 3 on pages 578 and 579, respectively. Hence no omission of cases arises through regarding the denominators in (6) as non-zero.

3. **The relations among the triangular area ratios.** We assume that in general no three of the points (x_i, y_i), $(i=1, 2, \cdots, 5)$, lie in the same straight line. Hence they uniquely determine a conic, and this conic may be projected into a circle. In this process of projection the triangular area ratios remain unchanged; that is, the λ's defined in (3) are invariant under projection. If we make use of this property, we find that when the D's are eliminated from (8), the λ's satisfy a system of equations of the form

$$\lambda_i - \lambda_{i+2}\lambda_{i+3} - \lambda_i\lambda_{i+2}\lambda_{i+3} = 0, \qquad i = 1, 2, \cdots, 5. \tag{9}$$

4. **A geometric property of the λ's.** Let us consider, as typical of the triangular area ratios, the first one, which by (8) is

$$\lambda_5 = D_{12}D_{34}/D_{14}D_{23}.$$

If the points 1, 2, 3, and 4 are assigned coordinates and λ_5 is assigned a value, then this equation represents a conic whose variables are the coordinates of the point 5, the nature of the conic depending upon the value given to λ_5. The conic obviously passes through the points 1, 2, 3, and 4. It will be useful later to know under what condition the conic will degenerate into a pair of straight lines.

Let the points 1, 2, 3, and 4 have, respectively, the coordinates (0, 0), $(x_2, 0)$, (x_3, y_3), and (x_4, y_4), let the coordinates of the point 5 be (x, y), and omit the subscript on λ_5. Then if the equation of the conic is written out and the condition for a conic to degenerate into a pair of straight lines is imposed, it is found that

$$\lambda(\lambda + 1) = 0.$$

Hence we have the following theorem:

THEOREM 4.1. *If, in any one of the triangular area ratios, the point indicated by the subscript on the associated* λ *is allowed to vary while the other four points are fixed, the variable point will move in a conic which passes through the four fixed points and will degenerate into a pair of straight lines if, and only if,* λ *has either the value* 0 *or* -1.

5. **The necessary condition.** In order that the problem may admit a solution other than the straight line solution, it is necessary that equations (7) be satisfied. However, if any three of these equations are satisfied, the remaining two are also satisfied, since any set of three of them implies the other two. For instance, suppose we write down the following three of these equations:

$$\begin{aligned} &S_{14}S_{23} - S_{13}S_{24}(1+\lambda_5) + S_{12}S_{34}\lambda_5 = 0 \\ &S_{15}S_{24} - S_{14}S_{25}(1+\lambda_3) + S_{12}S_{45}\lambda_3 = 0, \\ &S_{25}S_{34} - S_{24}S_{35}(1+\lambda_1) + S_{23}S_{45}\lambda_1 = 0, \end{aligned} \tag{10}$$

which correspond, respectively, to the values 1, 4, and 2 of i, and write down the one corresponding to $i=5$, namely,

$$S_{12}S_{35} - S_{13}S_{25}(1+\lambda_4) + S_{15}S_{23}\lambda_4 = 0. \tag{10a}$$

To show that this equation is implied by equations (10) multiply the second equation of (10) by $\lambda_1 S_{23}$ and the third one by $\lambda_3 S_{12}$ and subtract, thus eliminating S_{45}. Multiply this result throughout by λ_5 and the first equation of (10) by λ_3. Then, using the relation $\lambda_3 = \lambda_1\lambda_5 + \lambda_1\lambda_3\lambda_5$ from (9) and subtracting, we have

$$S_{13}S_{25}(1+\lambda_5) - S_{12}S_{35}(1+\lambda_1)\lambda_5 - S_{15}S_{23}(1-\lambda_1\lambda_5) = 0. \tag{11}$$

In order to see that this equation is identical with (10a), one only has to eliminate λ_2 from the second and fourth of the λ equations (9) obtaining $\lambda_4 = (1-\lambda_1\lambda_5)/(1+\lambda_1)\lambda_5$, and substitute this expression in (10a).

6. **Convex and concave pentagons.** Suppose that pegs are placed at each of the five masses and a string is drawn tautly around them. Two cases are distinguishable:

Case 1. The string touches all five pegs, and

(a) no three pegs are in a straight line;

(b) three pegs are in a straight line, but not four;

(c) four pegs are in a straight line, but not five;

(d) five pegs are in a straight line.

Case 2 (a). The string touches only four pegs, the fifth one being inside the quadrilateral formed by the other four.

Case 2 (b). The string touches only three pegs, the other two being inside the triangle formed by the first three.

If the conditions of Case 1 (a), 1 (b), or 1 (c) are satisfied, the pentagon will be called convex. Case 1 (d) is the straight line configuration with which we are not concerned. If the conditions of Case 2 (a) or Case 2 (b) are satisfied, the pentagon will be called concave.

In all cases we shall regard the masses as arranged in the counter-clockwise order m_1, m_2, m_3, m_4, and m_5.

From equations (6) we may write

$$
\begin{aligned}
m_1 &= \frac{D_{12}}{D_{23}} \frac{(R_{24}-R_0)(R_{35}-R_0)-(R_{34}-R_0)(R_{25}-R_0)}{(R_{14}-R_0)(R_{25}-R_0)-(R_{24}-R_0)(R_{15}-R_0)} m_3, \\
m_2 &= \frac{D_{12}}{D_{13}} \frac{(R_{34}-R_0)(R_{15}-R_0)-(R_{14}-R_0)(R_{35}-R_0)}{(R_{14}-R_0)(R_{25}-R_0)-(R_{15}-R_0)(R_{24}-R_0)} m_3, \\
m_2 &= \frac{D_{24}}{D_{34}} \frac{(R_{13}-R_0)(R_{45}-R_0)-(R_{14}-R_0)(R_{35}-R_0)}{(R_{14}-R_0)(R_{25}-R_0)-(R_{12}-R_0)(R_{45}-R_0)} m_3, \\
m_4 &= \frac{D_{24}}{D_{23}} \frac{(R_{12}-R_0)(R_{35}-R_0)-(R_{13}-R_0)(R_{25}-R_0)}{(R_{14}-R_0)(R_{25}-R_0)-(R_{12}-R_0)(R_{45}-R_0)} m_3, \\
m_4 &= \frac{D_{45}}{D_{35}} \frac{(R_{13}-R_0)(R_{25}-R_0)-(R_{23}-R_0)(R_{15}-R_0)}{(R_{24}-R_0)(R_{15}-R_0)-(R_{14}-R_0)(R_{25}-R_0)} m_3, \\
m_5 &= \frac{D_{45}}{D_{34}} \frac{(R_{14}-R_0)(R_{23}-R_0)-(R_{13}-R_0)(R_{24}-R_0)}{(R_{24}-R_0)(R_{15}-R_0)-(R_{14}-R_0)(R_{25}-R_0)} m_3, \\
m_1 &= \frac{D_{14}}{D_{45}} \frac{(R_{24}-R_0)(R_{35}-R_0)-(R_{34}-R_0)(R_{25}-R_0)}{(R_{12}-R_0)(R_{34}-R_0)-(R_{13}-R_0)(R_{24}-R_0)} m_5, \\
m_4 &= \frac{D_{14}}{D_{15}} \frac{(R_{13}-R_0)(R_{25}-R_0)-(R_{12}-R_0)(R_{35}-R_0)}{(R_{12}-R_0)(R_{34}-R_0)-(R_{13}-R_0)(R_{24}-R_0)} m_5, \\
m_1 &= \frac{D_{12}}{D_{25}} \frac{(R_{23}-R_0)(R_{45}-R_0)-(R_{24}-R_0)(R_{35}-R_0)}{(R_{13}-R_0)(R_{24}-R_0)-(R_{14}-R_0)(R_{23}-R_0)} m_5, \\
m_2 &= \frac{D_{12}}{D_{15}} \frac{(R_{14}-R_0)(R_{35}-R_0)-(R_{13}-R_0)(R_{45}-R_0)}{(R_{13}-R_0)(R_{24}-R_0)-(R_{14}-R_0)(R_{23}-R_0)} m_5.
\end{aligned}
\tag{12}
$$

For a convex pentagon, the ten triangular area ratios in these equations satisfy the inequalities

$$
\begin{aligned}
&D_{12}/D_{23}>0,\ D_{12}/D_{13}>0,\ D_{24}/D_{34}>0,\ D_{24}/D_{23}>0, \\
&D_{45}/D_{35}>0,\ D_{45}/D_{34}>0,\ D_{14}/D_{45}>0,\ D_{14}/D_{15}>0, \\
&D_{12}/D_{25}>0,\ D_{12}/D_{15}>0.
\end{aligned}
\tag{13}
$$

For the general concave pentagon of the type described in Case 2 (a), there are four different positions that the fifth mass may occupy within the quadrilateral formed by the other four masses. These positions are shown in Figs. 1a, 1b, 1c, and 1d. The following table gives the signs of the triangular area ratios for each of these four cases:

Figure	$\frac{D_{12}}{D_{23}}$	$\frac{D_{12}}{D_{13}}$	$\frac{D_{24}}{D_{34}}$	$\frac{D_{24}}{D_{23}}$	$\frac{D_{45}}{D_{35}}$	$\frac{D_{45}}{D_{34}}$	$\frac{D_{14}}{D_{45}}$	$\frac{D_{14}}{D_{15}}$	$\frac{D_{12}}{D_{25}}$	$\frac{D_{12}}{D_{15}}$
1a	−	+	+	+	+	+	+	+	−	−
1b	−	+	+	+	+	+	+	+	+	−
1c	−	−	+	+	+	+	+	+	+	−
1d	−	−	+	+	+	+	+	+	−	−

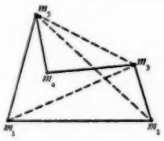


FIG. 1a

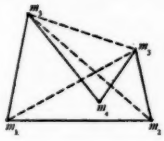


FIG. 1b

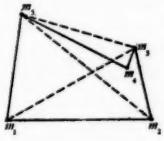


FIG. 1c

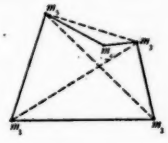


FIG. 1d

7. **Admissible convex pentagons.** An admissible pentagon is one for which there exist positive masses which, when placed at its vertices and started off suitably, will give a permanent configuration.

In any solution of the problem, one sees from (10) that the following equations

$$
\begin{aligned}
&\frac{(R_{14}-R_0)(R_{23}-R_0)-(R_{13}-R_0)(R_{24}-R_0)}{(R_{13}-R_0)(R_{24}-R_0)-(R_{12}-R_0)(R_{34}-R_0)}=\frac{D_{12}D_{34}}{D_{14}D_{23}},\\
&\frac{(R_{24}-R_0)(R_{15}-R_0)-(R_{14}-R_0)(R_{25}-R_0)}{(R_{14}-R_0)(R_{25}-R_0)-(R_{12}-R_0)(R_{45}-R_0)}=\frac{D_{12}D_{45}}{D_{24}D_{15}},\\
&\frac{(R_{34}-R_0)(R_{25}-R_0)-(R_{24}-R_0)(R_{35}-R_0)}{(R_{24}-R_0)(R_{35}-R_0)-(R_{23}-R_0)(R_{45}-R_0)}=\frac{D_{23}D_{45}}{D_{34}D_{25}}
\end{aligned}
\tag{14}
$$

must be satisfied, and in addition, for any solution, equations (12) must yield positive masses.

Since for convex pentagons the inequalities (13) must hold, and since the masses are necessarily positive, it follows that the numerator and denominator involving the R's in each of the equations (12) must be of like sign. If we let the numerator and denominator of the first equation be positive, the signs of the numerator and denominator in each of the other equations are determined, and we find that

$$(15)\qquad \begin{aligned}(R_{i,i+1} - R_0)(R_{i+2,i+3} - R_0) &> (R_{i,i+2} - R_0)(R_{i+1,i+3} - R_0)\\ &> (R_{i,i+3} - R_0)(R_{i+1,i+2} - R_0), \qquad i = 1, \cdots, 5.\end{aligned}$$

We can choose three of the points arbitrarily. Let these be the first three. We then have given r_{12}, r_{23}, and r_{13} and can therefore choose their order relation. Let this relation be

$$r_{12}, r_{23} \leqq r_0 \leqq r_{13}.$$

Draw r_{12} as in Fig. 2, and let m_1 and m_2 be at its extremities. With m_1 and m_2 as centers draw semicircles with radius r_0, intersecting at O. With O as center describe circular arc $a_1b_1c_4b_2a_2$ with radius r_0. Then with m_3 in the region Oa_2b_2, m_4 in the region bounded by circular arcs Ob_1, Ob_2, and $b_1c_4b_2$, and m_5 in the region Oa_1b_1, the inequalities (15) are all satisfied, subject to the condition

$$(16)\qquad r_{12}, r_{23}, r_{34}, r_{15}, r_{45} \leqq r_0 \leqq r_{13}, r_{14}, r_{24}, r_{25}, r_{35}$$

which assures that the masses in the regions specified are all positive. The proof that with the masses in these regions equations (14) are satisfied is too long to write down here. Only the method of procedure will be indicated.

In Fig. 2 take m_3 in its region as shown or on the boundary of this region, and with this point as center and with r_0 as radius describe circular arc $c_1c_2c_3c_4$, intersecting arc Oa_1 at c_2, arc Ob_1 at c_3, and arc b_1b_2 at c_4. On account of the inequalities (16), the mass m_4 is further restricted to the region $Oc_3c_4b_2$. Let this region be covered with an infinity of arcs, all passing through the point O and intersecting the boundary arc c_3c_4. It can be shown that on each of these arcs there is one and but one point at which the first of equations (14) is satisfied. If these points are joined by a curve, we have the curve on which point m_4 moves as m_3 varies its position in the region Oa_2b_2. Imagine m_4 on this curve, and with it as center and r_0 as radius describe circular arc a_3b_3 intersecting Ob_1 at b_3 and Oa_1 at some point a_3 below c_2. The remaining two equations of (14), then, determine two curves which pass through m_4 and intersect in the region $c_2c_3b_3a_3$, giving the position of m_5. As the point m_4 varies its position on its curve, m_5 moves on some curve in the region $c_2c_3b_3a_3$. These results lead to the following theorem:

THEOREM 7.1. *For every point m_3 in the region Ob_2a_2, or on the boundary of this region, there exists a single infinity of points m_4 in the region $Oc_3c_4b_2$ and a single infinity of points m_5 in the region Oa_1b_1, which together with the points m_1 and m_2 form an admissible convex pentagon.*

8. **Limitations on the interior angles.** In Fig. 2 let the angles of the pentagon at its five vertices be denoted by the masses at these points. For example, angle $m_2m_1m_5$ will be denoted by m_1, and angle $m_1m_2m_3$ by m_2.

It is obvious that angle m_1 is equal to or greater than angle m_2m_1O and that this latter angle obtains its least value when $r_{12}=r_0$. This value is seen to be 60° from Fig. 3, which is the same as Fig. 2 except that it is drawn with $r_0=r_{12}$. Therefore angle m_1 is equal to or greater than 60°, and the same is clearly true for angle m_2.

Since $r_{12}=r_0$, the first of equations (14) shows that the point m_4 moves on some curve through the point O, such as curve n, Fig. 3. We see, however, that the smallest value of angle m_3 is obtained when point m_3 is at b_2 and point m_4 is at O, in which case it is 60°. By symmetry it follows that the smallest value of angle m_5 is also 60° and is obtained when point m_4 is at O and point m_5 is at b_1.

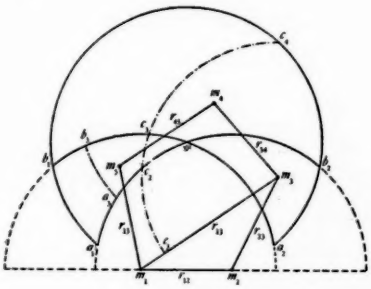


FIG. 2

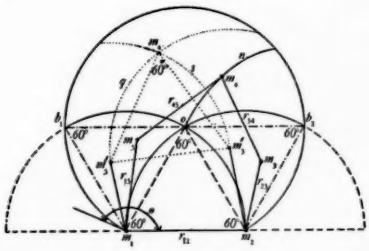


FIG. 3

Consider now angle m_4. In Fig. 3, let points m_3 and m_5 be at m_3' and m_5', respectively, so that r_{35} equals r_0, its least value. With these two points as centers describe arcs q and s, respectively, each with radius $r_0=r_{12}=r_{35}$. From the inequality (16) we see that the point m_4 cannot lie outside of the region bounded by the arcs q, s and the arcs of the circles drawn with the points m_1 and m_2 as centers. It is clear then that angle m_4 is smallest when point m_4 lies at the intersection of arcs q and s, and that it is then 60°.

Hence no interior angle of the convex pentagon can be less than 60°.

The maximum value of the angle m_1 is also reached when $r_{12}=r_0$, as can be seen from Figs. 2 and 3, and is that angle θ, Fig. 3, which the circle with O as center makes with the line r_{12}.

In order to find θ let the line through the points m_1 and m_2 be the x axis, and let these two points be symmetric with respect to the origin, so that the point O will be on the y axis. Then if the equation of the circle with O as center and r_0 as radius is formed, it follows easily that the angle θ which it makes with the line r_{12} is 150°.

It is obvious from Fig. 3 that angle m_2 also has this angle for its maximum value; and this is true for angles m_3, m_4, and m_5 as well, as can be shown by going through the same procedure for these angles as for angle m_1.

THEOREM 8.1. *Each of the interior angles of an admissible convex pentagon lies between* 60° *and* 150°.

9. **Admissible concave pentagons.** We shall consider the general case where one of the points m_i is within the quadrilateral formed by the other four. Let the vertices of the quadrilateral be m_1, m_2, m_3, and m_5, and let the interior point be m_4 whose position is that shown in Fig. 1d.

For concave pentagons of the type under consideration, the signs indicated in the last row in the table on page 569 must obtain, and since the masses are positive, the R's in equations (12) must satisfy the relations

$$
\begin{aligned}
&(R_{12}-R_0)(R_{45}-R_0) > (R_{14}-R_0)(R_{25}-R_0) < (R_{24}-R_0)(R_{15}-R_0),\\
&(R_{34}-R_0)(R_{15}-R_0) > (R_{14}-R_0)(R_{35}-R_0) > (R_{13}-R_0)(R_{45}-R_0),\\
&(R_{34}-R_0)(R_{25}-R_0) < (R_{24}-R_0)(R_{35}-R_0) > (R_{23}-R_0)(R_{45}-R_0),\\
&(R_{23}-R_0)(R_{15}-R_0) < (R_{13}-R_0)(R_{25}-R_0) > (R_{12}-R_0)(R_{35}-R_0),\\
&(R_{14}-R_0)(R_{23}-R_0) > (R_{13}-R_0)(R_{24}-R_0) < (R_{12}-R_0)(R_{34}-R_0).
\end{aligned}
\tag{17}
$$

If m_4 is in the position shown in either Fig. 1a or Fig. 1b, then the first inequality sign in the second inequality and the second inequality sign in the third inequality must be reversed, while there is no change if m_4 is in the position shown in Fig. 1c.

Here, as in the convex case, we may select three of the five points of the pentagon arbitrarily, and we shall take as these three the points m_1, m_2, and m_3. Also let the order relation between r_{12}, r_{23}, and r_{13} be $r_{12}, r_{23} \leqq r_0 \leqq r_{13}$.

In Fig. 4 draw r_{12} with the masses m_1 and m_2 at its extremities. With these two points as centers draw semicircles with radius r_0 intersecting at O. With O as center describe circular arcs a_1b_1 and a_2b_2, each with radius r_0. On account of the inequalities (17), m_3 will lie in the region Ob_2a_2, m_4 in Od_1d_2, and m_5 in Oa_1b_1, with the additional condition

$$r_{12}, r_{23}, r_{34}, r_{14}, r_{24}, r_{15}, r_{25}, r_{45} \leqq r_0 \leqq r_{13}, r_{25}.$$

With the masses in these regions it may be shown that equations (14) are satisfied. Hence we have the companion theorem:

THEOREM 9.1. *For every point m_3 in the region Ob_2a_2, or on the boundary of this region, there is a single infinity of points m_4 in the region Od_1d_2 and a single infinity of points m_5 in the region Oa_1b_1 which, together with the points m_1 and m_2, form an admissible concave pentagon.*

10. **Symmetric pentagons.** Suppose that

$$r_{23} = r_{15}, \qquad r_{14} = r_{24}, \qquad r_{34} = r_{45}, \qquad r_{13} = r_{25},$$

and consequently that

$$D_{23} = D_{15}, \qquad D_{45} = D_{34}, \qquad D_{14} = D_{24}, \qquad D_{25} = D_{13}.$$

The configurations Fig. 5 then present themselves in the form of isosceles trapezoids with vertices at the points m_1, m_2, m_3, and m_5 and with the point m_4 somewhere on the line perpendicular to and bisecting the parallel sides r_{12} and r_{35}. The construction of the figure is the same as that of Fig. 2.

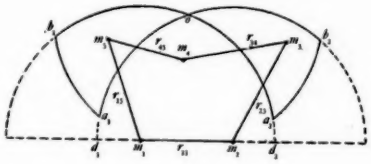


FIG. 4

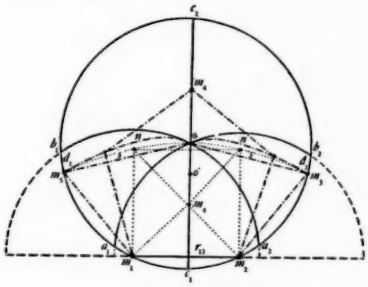


FIG. 5

Equations (14) reduce to the two independent equations

$$(18.1) \qquad \frac{(R_{24} - R_0)(R_{23} - R_{13})}{(R_{13} - R_0)(R_{24} - R_0) - (R_{12} - R_0)(R_{34} - R_0)} = \frac{D_{12}D_{45}}{D_{24}D_{15}},$$

$$(18.2) \qquad \frac{(R_{13} - R_0)(R_{34} - R_0) - (R_{24} - R_0)(R_{35} - R_0)}{(R_{24} - R_0)(R_{35} - R_0) - (R_{23} - R_0)(R_{34} - R_0)} = \frac{D_{15}}{D_{25}},$$

and the mass equations (12) become

$$(19.1) \qquad \frac{m_3}{m_1} = \frac{m_5}{m_2} = \frac{D_{15}}{D_{12}} \frac{(R_{24} - R_0)(R_{13} - R_{23})}{(R_{24} - R_0)(R_{35} - R_0) - (R_{13} - R_0)(R_{34} - R_0)},$$

$$(19.2) \qquad \frac{m_3}{m_2} = \frac{m_5}{m_1} = \frac{D_{25}}{D_{12}} \frac{(R_{24} - R_0)(R_{13} - R_{23})}{(R_{23} - R_0)(R_{34} - R_0) - (R_{24} - R_0)(R_{35} - R_0)},$$

$$(19.3) \qquad \frac{m_3}{m_4} = \frac{m_5}{m_4} = \frac{D_{15}}{D_{24}} \frac{(R_{12} - R_0)(R_{34} - R_0) - (R_{13} - R_0)(R_{24} - R_0)}{(R_{13} - R_0)(R_{13} - R_0) - (R_{12} - R_0)(R_{35} - R_0)},$$

$$m_5 = m_3.$$

From either (19.1) or (19.2) it follows that we have also $m_1 = m_2$.

For convex pentagons all of the triangular area ratios in these equations are positive. For concave pentagons D_{15}/D_{12} and D_{25}/D_{12} are negative, if the point m_4 is above the diagonal r_{13}; otherwise they are positive. Moreover $D_{15}/D_{24}>0$.

The convex case. In Fig. 5, r_{12} is regarded as fixed with the masses m_1 and m_2 at its extremities, while the point m_3 (and its symmetric point m_5) varies when the point m_4 moves from O towards c_2, as m_4 must do since $r_{24}\geqq r_0$ for convex pentagons.

When the point m_4 is at O, equations (18.1) and (18.2) become, respectively,

$$(R_{12}-R_0)(R_{34}-R_0)D_{12}D_{45}=0,$$

$$(R_{34}-R_0)\left[(R_{13}-R_0)D_{25}+(R_{23}-R_0)D_{15}\right]=0.$$

In the first equation since neither $(R_{12}-R_0)$, D_{12}, nor D_{45} is zero, we must have $(R_{34}-R_0)=0$; whence $r_{34}=r_0$. It then follows that point m_3 is somewhere on the circular arc $m_1b_1c_2b_2a_2m_2$, and it can be proved that it is at some point d_2 between a_2 and b_2 with m_5 at the symmetric position d_1.

As the point m_4 moves from O towards c_2, point m_3 moves on some curve n from d_2 towards O, and in like manner, point m_5 moves from d_1 towards O along some curve n' which is symmetric to n. When m_4 reaches a position on Oc_2 such that the points m_3 and m_5 have moved in towards each other on curves n and n', respectively, so that $r_{35}=r_0$, then the limit for permanent convex configurations is attained, since by inequalities (16), $r_{35}\geqq r_0$.

The concave case. As the point m_4 moves from O towards c_1, as it must do for concave pentagons since $r_{24}\leqq r_0$, point m_3 moves on some curve s from d_2 towards O, and in like manner, point m_5 moves on some curve s', symmetric to s from d_1 towards O. However, since for concave pentagons $r_{35}\leqq r_0$, these curves are of no interest until points m_3 and m_5 reach positions on curves s and s', respectively, where $r_{35}=r_0$.

Let the corresponding position of m_4 be O'. If one starts with the point m_4 in this position, as it moves on towards c_1 points m_3 and m_5 move towards O. When they reach this position one has $r_{13}=r_{23}=r_0$, $D_{12}=0$, $D_{15}/D_{25}=-1$. Equations (18.1) and (18.2) are then both satisfied, regardless of the position of point m_4 on Oc_1.

Thus, for a given r_{12} and r_0, as point m_4 moves from O towards c_2, point m_3 moves from d_2 towards O along n, and point m_5 moves from d_1 towards O along n', giving, in all positions for which $r_{35}\geqq r_0$, a convex pentagon; while as point m_4 moves from O' towards c_1, points m_3 and m_5 move towards O along s and s', respectively, giving, in all positions where $r_{35}\leqq r_0$, a concave pentagon. Since the mass ratios are all positive for all of these pentagons, as

can be seen from equations (19.1), (19.2), and (19.3), we can state the following theorem:

THEOREM 10.1. *For every r_{12} and r_0 there exists a single infinity of permanent configurations with one mass at each of the vertices of an isosceles trapezoid and with the fifth on the line perpendicular to and bisecting the two parallel sides.*

As the point m_4 moves from O towards c_2, Fig. 5, and consequently, points m_3 and m_5 move from d_2 and d_1 towards the symmetric points on n and n' where $(r_{35}-r_0)$ vanishes, one sees from equations (19.1), (19.2), and (19.3) that the mass ratios

$$\frac{m_3}{m_1}=\frac{m_5}{m_2}, \qquad \frac{m_3}{m_2}=\frac{m_5}{m_1}, \qquad \frac{m_3}{m_4}=\frac{m_5}{m_4},$$

change steadily from the values

$$\frac{D_{15}(R_{13}-R_{23})}{D_{12}(R_{35}+R_{13}-2R_0)}, \qquad \frac{D_{25}(R_{13}-R_{23})}{D_{12}(2R_0-R_{23}-R_{35})}, \qquad 0,$$

respectively, to

$$\frac{D_{15}(R_{24}-R_0)(R_{13}-R_{23})}{D_{12}(R_0-R_{13})(R_{34}-R_0)}, \qquad \frac{D_{25}(R_{24}-R_0)(R_{13}-R_{23})}{D_{12}(R_{23}-R_0)(R_{34}-R_0)},$$

$$\frac{D_{15}(R_{12}-R_0)(R_{34}-R_0)-(R_{13}-R_0)(R_{24}-R_0)}{D_{24}(R_{13}-R_0)^2},$$

where the D's and R's are to be evaluated at the points in question. Similar limits on the mass ratio could be found for the concave case as m_4 moves from O' towards c_1. We therefore have the following theorem:

THEOREM 10.2. *For every $m_1=m_2>0$, $m_3=m_5>0$, and $m_4>0$, there is one and only one isosceles trapezoid configuration.*

11. **The case of three points on a line.** Let m_4 be on the line r_{13} joining m_1 and m_3, and let m_2 and m_5 be equidistant from r_{13}. We have at once

$$\begin{aligned} &r_{12}=r_{15}, \qquad r_{23}=r_{35}, \qquad r_{24}=r_{45}, \\ &D_{25}=0, \qquad D_{35}=D_{23}, \qquad D_{45}=D_{24}, \qquad D_{15}=-D_{12}, \end{aligned} \tag{20}$$

and the equations (14) reduce to the single equation

$$\frac{(R_{14}-R_0)(R_{23}-R_0)-(R_{13}-R_0)(R_{24}-R_0)}{(R_{13}-R_0)(R_{24}-R_0)-(R_{12}-R_0)(R_{34}-R_0)}=\frac{-D_{34}D_{15}}{D_{14}D_{35}}. \tag{21}$$

Let points m_1 and m_3 be fixed. Then for a given m_2 and for any position of point m_4, the equation (21) to be satisfied, by the coordinates of point m_5,

is the equation of a straight line. But for a solution the coordinates of m_5 must also satisfy the first two equations of (20). (The remaining equations of (20) are then automatically satisfied.) These equations, when rationalized, represent circles with centers at the points m_1 and m_3 and with radii r_{12} and r_{23}, respectively.

In Fig. 6 draw r_{13} with m_1 and m_3 at its extremities. Take point m_2 anywhere not on this line, and draw the circles of the above equations. It is obvious that these two equations are satisfied by real points if, and only if, point m_5 lies at the intersection of their circles. That is, point m_5 will either be at the point a, Fig. 6, or it will coincide with point m_2. Thus it may be in one of two positions in so far as these two equations are concerned.

Equation (21) must be satisfied also. By Theorem 4.1, the left member of this equation must be either 0 or -1. If it is 0, the equation reduces to

$$D_{34} = 0,$$

which is the equation of a straight line passing through points m_1 and m_2; therefore, in order for the point m_5 to satisfy all the equations, it must coincide with m_2. This is a case of no interest as r_{25} may vanish for neither convex nor concave pentagons.

Suppose the left-hand member has the value -1. This gives the equation

$$\frac{R_{12} - R_0}{R_{23} - R_0} = \frac{R_{14} - R_0}{R_{34} - R_0}. \tag{22}$$

The question arises as to whether or not the line with this equation can be made to pass through the point a. In other words, are there configurations other than the one where points m_2 and m_5 coincide? The proof that the answer to this question is in the affirmative is not difficult, but an illustrative example would, perhaps, be more instructive.

Let the point m_4 be midway between points m_1 and m_3. The equation (22) is always satisfied, and if the point m_2 is given a position such that the mass ratios are all positive, we have a permanent configuration which, in general, will be a rhombus with vertices at m_1, m_2, m_3, and m_5, and with m_4 at its center.

In Fig. 7, draw r_{13}, and consider the point m_2 as it moves on a line perpendicular to r_{13} at its mid-point m_4. The positions that it may assume are limited only by the regions for positive masses, since the condition (22) is satisfied for any position of m_2 regardless of what r_0 may be. Moreover wherever point m_2 may be on this line, we have, from (12),

$$\frac{m_1}{m_3} = 1, \qquad \frac{m_2}{m_5} = 1.$$

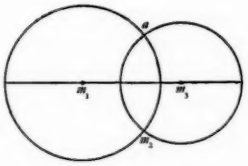


FIG. 6

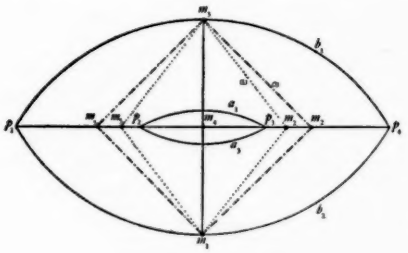


FIG. 7

With points m_1 and m_3 as centers draw circular arcs a_1 and a_3 with a radius r_0, $(r_0 < r_{13})$, such that the distance between their intersections p_2 and p_3 is equal to r_0. Since $r_{25} \geqq r_0$, we have, with point m_2 at p_3 and m_5 at p_2, the rhombus configuration in which the point m_2 is as close to m_5 as possible for a given r_{13}. In this case the mass equations (12) show that

$$(23) \quad \frac{m_3}{m_2} = 0, \quad \frac{m_4}{m_3} = \frac{2(R_0 - R_{13})}{(R_{14} - R_0)} > 0, \quad \frac{m_3}{m_5} = 0, \quad \frac{m_1}{m_5} = 0, \quad \frac{m_4}{m_5} = 0.$$

The greatest possible distance between the points m_2 and m_5, for a given r_{13}, is obtained when we choose an $r_0 = r_{13}$. Draw circular arcs b_1 and b_2 with such a radius. Then, with m_2 and p_4 and m_5 at p_1, we have the rhombus configuration in which the distance between points m_2 and m_5 is as great as possible. Here, we have

$$(24) \quad \frac{m_3}{m_2} = \infty, \quad \frac{m_4}{m_3} = 0, \quad \frac{m_3}{m_5} = \infty, \quad \frac{m_1}{m_5} = \infty, \quad \frac{m_4}{m_5} = \frac{2(R_0 - R_{25})}{(R_{24} - R_0)} > 0.$$

THEOREM 11.1. *Given r_{13} with masses m_1 and m_3 at its extremities and with the mass m_4 at its center, for a properly chosen r_0, and for all positions of m_2 and m_5 on a line perpendicular to r_{13} at m_4 such that the mass ratios have values lying between those given in* (23) *and those in* (24), *the configuration is permanent and is a rhombus with vertices at m_1, m_2, m_3, and m_5, and with m_4 at its center.*

In concluding this section we may add that if any four of the masses are on the same straight line, the fifth mass will be on this line also. This follows readily from equations (14).

12. **The triangle with two interior masses.** One other possibility of a concave pentagon is that of a triangle with one mass at each vertex and with the two remaining masses in its interior. That no such configuration, however, can form an admissible pentagon follows from the inequalities

$$r_{12}, r_{23}, r_{34}, r_{14}, r_{24}, r_{15}, r_{35}, r_{45} \leqq r_0 \leqq r_{13}, r_{25}$$

(see §9). To construct a pentagon of this type one must alter the above order relation that must exist among the sides and diagonals, and the result is always to cause some of the masses to be negative.

13. **Numerical solutions.** We consider the following examples:

EXAMPLE 1. An interesting example of a convex permanent configuration is that of the regular pentagon. In Fig. 5 as the point m_4 moves upward from O towards c_2 and the points m_3 and m_5 move towards each other along curves n and n', somewhere along these paths the configuration becomes that of a regular pentagon. When this position is reached, the equations (18.1) and (18.2) reduce to the single equation

$$\frac{R_{24} - R_0}{2R_0 - R_{12} - R_{24}} = \frac{D_{15}}{D_{25}} = \frac{r_{12}}{r_{24}}, \tag{25}$$

or

$$(r_0^3 - r_{24}^3)r_{12}^2 r_{24} - 2r_{12}^3 r_{24}^3 + r_0^3(r_{24}^3 + r_{12}^3) = 0. \tag{26}$$

For any given value of r_{13}, r_{24} may be computed from the obvious relation

$$r_{24} = 1.6180 r_{12},$$

and r_0 may be found from (26). It follows readily that

$$\frac{r_0}{r_{12}} = 1.3076, \qquad \frac{r_{23}}{r_{12}} = \frac{r_{34}}{r_{12}} = \frac{r_{45}}{r_{12}} = \frac{r_{15}}{r_{12}} = 1,$$

$$\frac{r_{13}}{r_{12}} = \frac{r_{24}}{r_{12}} = \frac{r_{14}}{r_{12}} = \frac{r_{25}}{r_{12}} = \frac{r_{35}}{r_{12}} = 1.6180,$$

and from (19.1), (19.2), (19.3), together with (25), that $m_1 = m_2 = m_3 = m_4 = m_5$.

EXAMPLE 2. Perhaps the most simple concave permanent configuration in the five-body problem is that of a square with one mass at each of its vertices and with the fifth mass at its center. In Fig. 5 this particular configuration is passed through as the point m_4 moves from O towards c_1 and as points m_3 and m_5 move towards each other along curves s and s'. Equations (18.1) and (18.2) are now identically satisfied for any value of r_0, while the mass equation (19.1) shows that

$$\frac{m_3}{m_1} = \frac{m_5}{m_2} = 1;$$

and since for all of the symmetric configurations, Fig. 5, $m_3 = m_5$, $m_1 = m_2$, we have $m_1 = m_2 = m_3 = m_5$.

The equation (19.3) becomes

$$\frac{m_3}{m_4}=\frac{m_5}{m_4}=\frac{1}{2}\,\frac{R_{34}-R_0}{2R_0-R_{12}-R_{13}}=\frac{(r_0^3-r_{34}^3)r_{12}^3r_{13}^3}{2r_{34}^3(2r_{12}^3r_{13}^3-r_0^3r_{13}^3-r_0^3r_{12}^3)}$$

(27)

$$=\frac{(8r_0^3-2(2)^{1/2}r_{12}^3)}{2[4(2)^{1/2}r_{12}^3-r_0^3(1+2(2)^{1/2})]},$$

since $r_{13}^2=2r_{12}^2=4r_{34}^2$.

Thus, while equations (18.1) and (18.2) are satisfied for any r_0, the mass ratios $m_3/m_4=m_5/m_4$ are not necessarily positive for any value of r_0. It is found that for the last right-hand member of (27) to be positive we must have

$$0.878<\frac{r_{12}}{r_0}\leqq 1.000.$$

It then follows that

$$0.439<\frac{m_3}{m_4}=\frac{m_5}{m_4}\leqq 1.414.$$

This result shows that the central mass is arbitrary within certain limits. In general, if a permanent configuration is given, the masses are uniquely determined.

EXAMPLE 3. In Fig. 7 consider the rhombus configuration (1) with masses m_1, m_2, m_3, and m_5 at its vertices and with the mass m_4 at its center. By Theorem 11.1, for a given r_0 the point m_2 may be anywhere on the line perpendicular to r_{13} at its mid-point such that the mass ratios are positive. It is readily found that

$$\frac{r_0}{r_{12}}=1.042,\qquad \frac{r_{23}}{r_{12}}=\frac{r_{35}}{r_{12}}=\frac{r_{15}}{r_{12}}=1,\qquad \frac{r_{14}}{r_{12}}=\frac{r_{34}}{r_{12}}=0.781,$$

$$\frac{r_{24}}{r_{12}}=\frac{r_{45}}{r_{12}}=0.625,\qquad \frac{r_{25}}{r_{12}}=1.249,\qquad \frac{r_{13}}{r_{12}}=1.562;$$

and from the mass equations (12) we have,

$$\frac{m_1}{m_3}=1,\quad \frac{m_2}{m_3}=\frac{m_5}{m_3}=\frac{m_5}{m_1}=2.617,\quad \frac{m_4}{m_3}=0.536,\quad \frac{m_5}{m_4}=4.878,\quad \frac{m_2}{m_5}=1.$$

As the point m_2 moves on towards p_4 and the point m_5 moves towards p_1, Fig. 7, the polygon moves through the square configuration (2), but this has already been discussed in Example 2 in connection with Fig. 5.

THE UNIVERSITY OF CHICAGO,
CHICAGO, ILL.